Recognized PoC Report – Significant Step Forward

Inter-DC VM Migration and Long-Distance DB Replication for Financial Industry

DECEMBER 2025

Contents

1 Introduction

This document serves as the Proof of Concept (PoC) report for the Service Infrastructure for Financial Industry Use Case Project. It provides a thorough analysis and evaluation of the implemented PoC, assessing its alignment with the Reference Implementation Model and Proof of Concept Reference guidelines [FSI Use Case RIM PoC Ref]. The report aims to verify the feasibility, effectiveness, and scalability of the proposed infrastructure solutions in addressing the unique challenges faced by the financial industry.

The Services Infrastructure for the Financial Industry Use Case Project aims to enhance performance and reliability with agile and resilient services. Using the IOWN Global Forum Open All-Photonic Networks, it provides energy-efficient, high-performance systems and seamless workload migration, lowering costs and increasing ROI across competitive financial markets.

The scope of the project includes:

- Documenting the Services Infrastructure for the Financial Industry Use Case and its key requirements.
- Establishing technology evaluation criteria, including reference cases and critical benchmarks.
- Creating the Reference Implementation Model as a practical guide for integrating the IOWN Global Forum technologies to realize the use case.
- Defining the Proof of Concept (PoC) Reference, providing structured guidelines for the PoC to assess the Reference Implementation Model using the Technology Evaluation Criteria.
- Developing and evaluating the PoC based on the Reference Implementation Model and PoC Reference.

This document follows on from the [FSI Use Case] and [FSI Use Case RIM PoC Ref] for the Financial Industry Use Case, and describes the details of the PoC conducted

We have concluded the two PoCs as mentioned below:

PoC 1: Intra-regional use case - VM migration across multiple data centers
PoC 2: Inter-regional use case - Long-distance DB sync/async replication

2 PoC project overview

2.1 PoC scope and outcomes

The Proof of Concept (PoC) aims to validate the applicability and effectiveness of the IOWN Global Forum technologies in the financial services infrastructure. The PoC scope includes evaluating VM migration performance, data replication performance, and disaster recovery performance as a whole across regions and data centers. The expected outcomes are to demonstrate the feasibility of instant and robust disaster recovery using the forum technologies, even when the daily data processing throughput is constantly high.

Hypothesis and the proved outcome of the PoC

1. The IOWN Global Forum enables financial systems to move flexibly among distributed data centers. The PoC demonstrates the feasibility of such a flexible system movement.
2. The IOWN Global Forum allows financial systems to back up in real time. The PoC resulted in an acceptable performance regression against the real-time backup scenario.

Core goals of the PoC

The primary goals of the PoC are:

- VM migration can be performed with minimal downtime to demonstrate resource allocation optimization, ensuring business continuity.
- The recovery point objective (RPO) and Recovery Time Objective (RTO) for different system tiers must be measured to validate the disaster recovery capabilities.
- To evaluate the performance of synchronous and asynchronous replication methods regarding latency, throughput, and data integrity.

Summary of PoC Reference Implementation and Evaluation

The correspondence between each PoC and the [FSI Use Case] and [FSI Use Case RIM PoC Ref], as well as the PoC results, is described below.

Table 2-1: Summary of PoC results

	Use Case	Key Requirement (KR)	Validation Point	Result	Relevant Section
PoC 1	Intra-regional use case	KR1.1	Downtime ≤ 1[s]	Pass	5.1.4
PoC 2	Inter-regional use case	KR2.1	RPO=0[*1]	Pass[*5]	5.2.4
		KR2.2	• Tier 1: RTO ≤ 30[min][*2] • Tier 2: RTO ≤ 2[h][*3] • Tier 3: RTO ≤ 24[h][*4]	Pass[*6]	5.2.4

Note 1: RPO is not intended to be achieved solely by the DBMS but also includes the application layer and the manual operations of the financial company. [FSI Use Case]
Note 2: Additional time is acceptable if a human operator needs to restore unprocessed transaction data or check the consistency of data. [FSI Use Case]
Note 3: The backup period is not always real-time, and hourly or daily is ordinary timing for a full backup. [FSI Use Case]
Note 4: The backup period is not always real-time, and daily or special maintenance time (e.g., holidays) is also ordinarily a time for full backup. [FSI Use Case]
Note 5: Synchronous replication was confirmed feasible at all distances defined in the Inter Region use case. Therefore, RPO=0 can be considered achieved solely at the DBMS layer.
Note 6: Synchronous replication was confirmed feasible at all distances defined in the Inter Region use case. Therefore, RTO=0 can be considered achieved solely at the DBMS layer.

PoC 1: Intra-regional use case - VM migration across multiple data centers

In this PoC, we successfully met the strict requirement of less than one second of downtime, as specified in [FSI Use Case RIM PoC Ref]. Generally, a threshold of one second is considered imperceptible to end-users in client-server systems with human interaction, so it does not suggest a system failure or malfunction.

This result suggests the potential for more adaptable deployment of application components on the server side without compromising user experience. Additionally, it demonstrates a possibility of sharing ICT resources across data centers, thereby confirming the practicality of dynamic resource management in intra-regional infrastructures.

PoC 2: Inter-regional use case - Long-distance DB sync/async replication

Traditionally, synchronous database replication has been considered unsuitable for long-distance environments over approximately 100 km and is often cited as a common anti-pattern in system design. As a result, in scenarios requiring remote data redundancy—such as active-standby system architectures—async replication or periodic backup strategies have been the standard approach.
In our PoC, it was shown that, with 300 concurrent connections, synchronous replication could be completed in less than 10 milliseconds from the application's perspective, even over approximately 500 km. APN extends the distance over which synchronous replication can be performed while meeting performance requirements.

It should be noted, however, that the acceptable replication latency perceived by the application may vary based on each system's specific requirements. Therefore, our PoC results should serve as a reference for system architects and engineers considering the adoption of synchronous replication in real-world deployments, especially when regional separation between primary and secondary systems is desired.

2.2 PoC completion status

Table 2-2: PoC project name and status

PoC Project Name	Inter-DC VM Migration and Long-Distance DB Replication for Financial Industry
PoC Completion Status	Completed
PoC Stage Completion Status	Completed

2.3 PoC project participants

Table 2-3: PoC project participants

Company	**Name(s)**
MUFG Bank, Ltd.	Masayoshi Namba and Suguru Yasui, et al.
NTT, Inc and Group companies	Naoto Umemori, Kimihiko Suga, Koyo Yoshida, Masanobu Yoshi and Yusuke Suzuki, et al.

3 Supported use case

3.1 Intra-regional application deployment and migration

Description

As illustrated in Figure 3-1, this use case enhances computing resource usage with resource pooling and dynamic allocation for resiliency, agility, and cost optimization. Seamless application migration boosts scalability and supports services like Banking-as-a-Service, which face unpredictable transaction volumes. Resiliency measures include relocating low-priority services for topical events or evacuating all services during local disasters like fires. A financial institution integrates private data centers, co-location spaces, offices, and private/public clouds into one logical computing zone within the same region.

Key requirements

The computing zone must support VM migration with downtime under 1 second to ensure smooth operation. Migration involves copying VM data, transferring changes, suspending, and resuming VMs on new servers. Downtime ensures consistent data and configurations, allowing applications to function seamlessly post-migration.

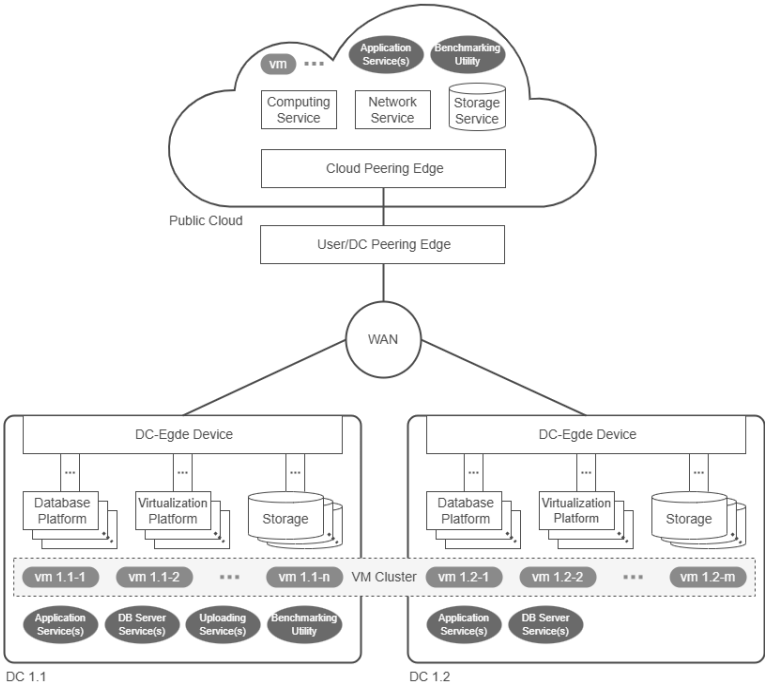

Figure 3-1: Reference system architecture for intra-regional use case

3.2 Inter-regional back-ups and migration

Description

As illustrated in Figure 3-2, this use case enables disaster recovery and inter-regional workload migration for energy balancing by utilizing multi-region infrastructures with near real-time data replication. The IOWN Global Forum technology can simplify systems and enhance their effectiveness. For example, it enables the placement of the secondary system directly at the backup data center, allowing us to bypass the redundant intermediate system typically found in the main data center. This approach not only streamlines the system architecture but also significantly improves data security and system efficiency. Geographical distances between sites are defined for disaster measures, with RPO/RTO prioritized for business continuity. There are priorities in system groups because business continuity is one of the key demands of today's financial organizations. A financial institution has the highest priority for business-critical data and services, such as accounting systems and its ledger data (Tier 1 Systems, Top Priority Group, < 1% of all systems). It also places major systems as a moderate priority to be resumed in a reasonable duration (Tier 2 Systems, Second Priority Group, < 10% of all systems). Data synchronization and application migration are mandatory to realize this system's autonomous recovery. The remaining systems, other than Tier 1 and Tier 2, the general group, are referred to as Tier 3.

Key requirements

Without losing user data, financial systems must achieve zero RPO across Tiers 1-3. Synchronous replication protects the database but reduces update throughput, with latency worsening performance. Asynchronous replication reduces this risk but can leave updates unreplicated in a disaster. A balanced solution is needed for throughput and recovery. RTO requirements are Tier 1: <30 minutes, allowing manual transaction restoration; Tier 2: <2 hours, typically hourly/daily backups; Tier 3: <24 hours, using daily backups or special maintenance periods like holidays.

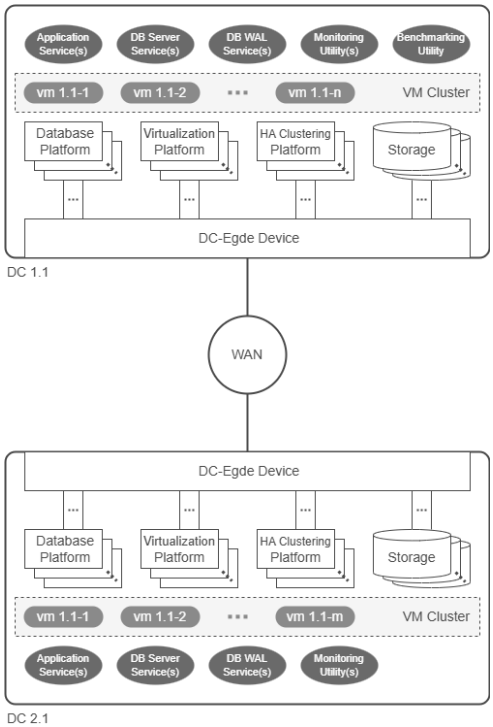

Figure 3-2: Reference system architecture for inter-regional use case

4 PoC system configuration and implementation

This section describes the PoC system configuration and implementation for each use case.

4.1 PoC 1: VM migration across multiple DCs

The PoC targets deploying virtual machine (VM) clusters within a two-data-center configuration, enabling system migration from one data center to another. The basic configuration is implemented to establish connectivity and meet the requirements in the reference implementation model [FSI Use Case RIM PoC Ref].

4.1.1 PoC 1: System overview

In the Financial Industry Use Case, the required network bandwidth for each application server is defined as 25 Gbps. Among the four IA servers shown in Figure 4-1, three are equipped with two 25 GbE NICs and are used as application servers. Of the two NICs, one is dedicated to virtualized storage traffic, while the other is used for all remaining communication purposes. These different types of traffic are logically isolated through network isolation features to ensure proper separation of communication paths.

The IA server equipped with a 100 GbE NIC is used to generate load (background traffic) on the Open APN. The Leaf Switch functions as a typical Top-of-Rack (ToR) Layer 2 switch, while the Leaf SW operates as a typical Layer 3 switch. The client-side interface of the APN-T supports 100 GbE. In PoC 1, the APN-T line-side interface outputs two wavelengths, each of which is used for different purposes. For details on servers used in PoC 1, please refer to Table B-1 in Appendix B.

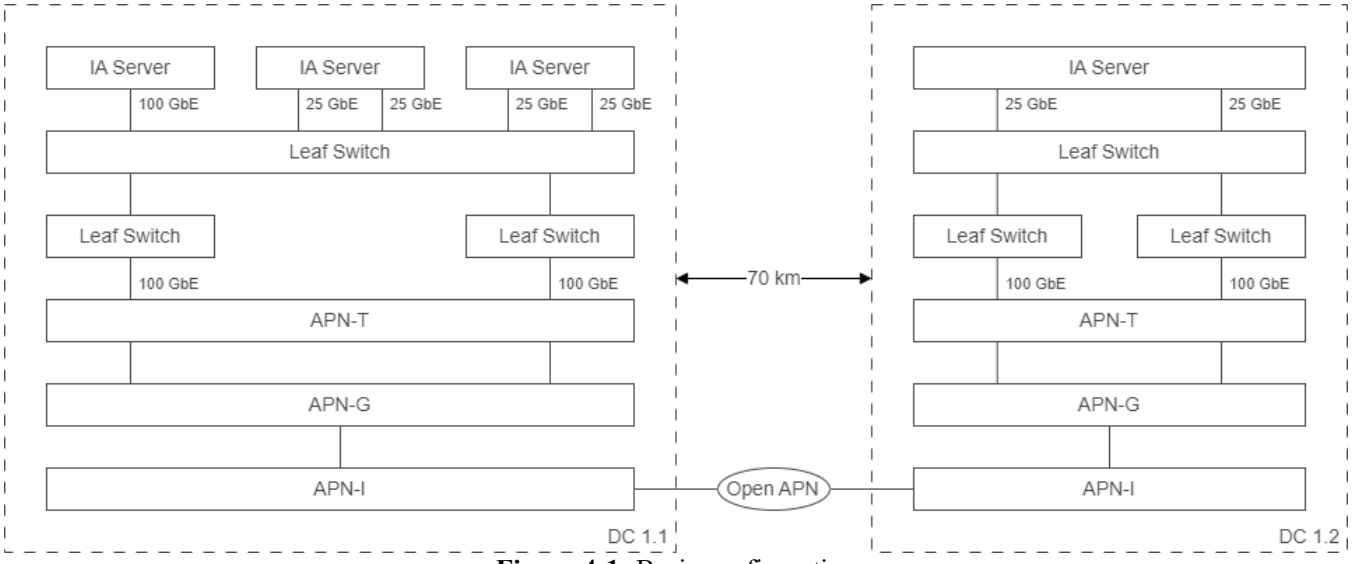

Figure 4-1: Basic configuration

4.1.2 PoC 1: PoC implementation

In PoC 1, a virtualized infrastructure is deployed across data centers interconnected via the Open APN. This virtualized platform hosts a pseudo-online banking system consisting of distributed storage and a Web 3-tier architecture configured in an active-active (ACT/ACT) setup.

The following sections describe the implementation policies and details for each layer: the software, platform, and network infrastructure layers.

Software layer

Figure 4-2 and Table 4-1 show that the pseudo-online banking system comprises multiple virtual machines (VMs) and applications.

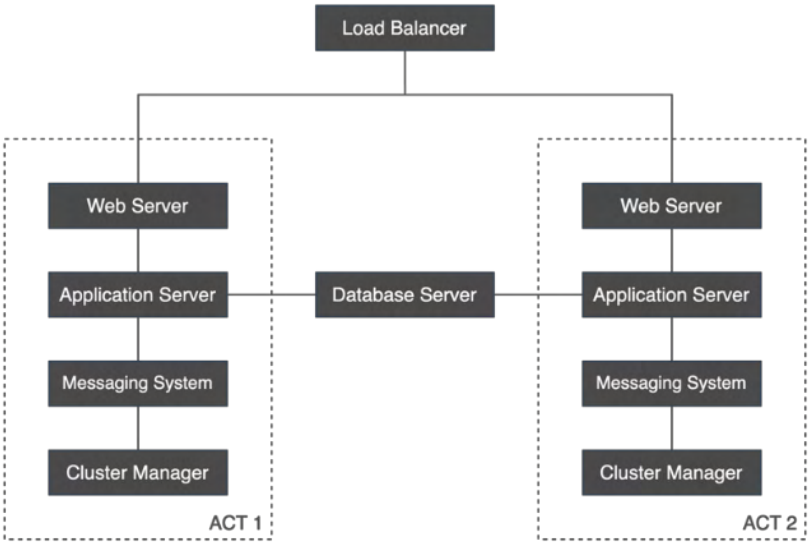

Figure 4-2: Configuration of pseudo-online banking application system

Table 4-1: OS and software list for PoC 1

Components	Software	Version
OS	Rocky Linux	9.3
Virtualization Platform	Docker	26.0.1
	VMware vCenter Foundation	8.0
	VMware vSAN Advanced	8.0
	VMware vSphere Standard	8.0 update 2
Load Balancer	HAProxy	3.0.0
Web Server	Nginx	1.27.0
Application Server	Java (OpenJDK)	11.0.22
	OnlineBankingRestAPI	(commit id: bd9d6fce91d0fe80a7abbbeac173f37d540344b0)
	Spring Boot	2.6.5
Database Server	PostgreSQL	16.3
Messaging System	Apache Kafka	2.8.1
Cluster Manager	Apache ZooKeeper	3.4.13
Load Generator	Apache JMeter	5.6.3
	fio	3.35
	iperf	3.16

Application layer

The verification application used in PoC 1 provides a RESTful API. The operation history can be stored in a DB when each API is executed. A registered user's login authentication (Token) is required to perform each API. The token obtained has a validity period of 24 hours. The verification application does not implement a function to prevent double-checking operations.

Table 4-2: OnlineBankingRestAPI resources related to this PoC

Resource	Operation
Customer	Save, update, delete, find all customers and find customers by id
Account	CRUD operations, also transfer money (transferring money between accounts), withdraw (getting money from an account), and deposit (loading money into the account)
AccountActivity	Action histories are saved for withdraw and deposit
MoneyTransfer	Action histories are saved in case of transfer
CreditCard	Create credit cards and do CRUD operations
CreditCardActivity	Action histories are saved regarding credit card activities
CreditCardActivityAnalysis	Learn the min, max, average, and count of credit card activities grouped by activity type (spend, refund, receive)
Loan	Calculate the loan by calculating the loan endpoint by the installment count and the principal money amount given by the customer

Table 4-3 shows the table definitions used in the verification application.

Table 4-3: Database tables defined for the OnlineBankingRestAPI resources in Table 4-2

Table	Purpose
acc_account	Manage account information
acc_account_activity	Record information on deposits, withdrawals, transfers, and receipts
acc_money_transfer	Record money transfers used between accounts
crd_credit_crd	Manage credit card information
crd_credit_crd_activity	Keep track of money transfers spent on credit cards
cus_customer	Manage customer information
loa_loan	Manage loan information
loa_loan_payment	Keep track of loan payments
log_detail	Record detailed information in the operation log

Platform layer

In PoC 1, live migration of virtual machines (VMs) is performed. The downtime associated with live migration depends on the specifications of the VMs. In this PoC, the VM specifications comply with those defined in Section 3.1.1 of [FSI Use Case].

- Virtual CPU: 2 Cores
- RAM: 16 GB
- Disk: 100 GB (Thick Provisioning)

Network infrastructure layer

In PoC 1, the VXLAN over an IP network is combined with the Open APN mechanism. For specific VXLANs (traffic identified by VNIDs), traffic is steered through designated Optical channel Data Unit (ODU) paths to extend Layer 2 connectivity.

PoC 1 utilizes two ODU paths. The type of traffic assigned to each ODU path varies depending on the validation scenario.

The conditions for the load balancer to be used in PoC 1 are as follows:

- Balancing Mode: Round-robin
- HTTP Timeout: 1 [sec]
- The number of concurrent access limitations: 3,000
- The number of connection pools for Application Server: 10
- The number of connection pools for Database Server: 100

4.2 PoC 2: Long-distance DB sync/async replication

The PoC targets verify inter-regional backups by performing synchronous and asynchronous data replication between database servers located far apart. The verification is conducted based on the measurement indicators defined in the reference implementation model [FSI Use Case RIM PoC Ref].

4.2.1 PoC 2: System overview

The PoC was conducted using the configuration shown in Figure 4-3. In this setup, PostgreSQL is running on both the DB Server (primary) and DB Server (secondary), and each server stores data on SAN Storage connected via Fibre Channel. To generate transactions, we used pgbench, a benchmark tool included with PostgreSQL.

Since preparing physical optical fibers with lengths ranging from 250 to 2500 km was challenging, we adjusted the Network Emulator to simulate the latency equivalent to using optical fibers of those lengths. Further details about servers and SAN storage are provided in Appendix B.2.

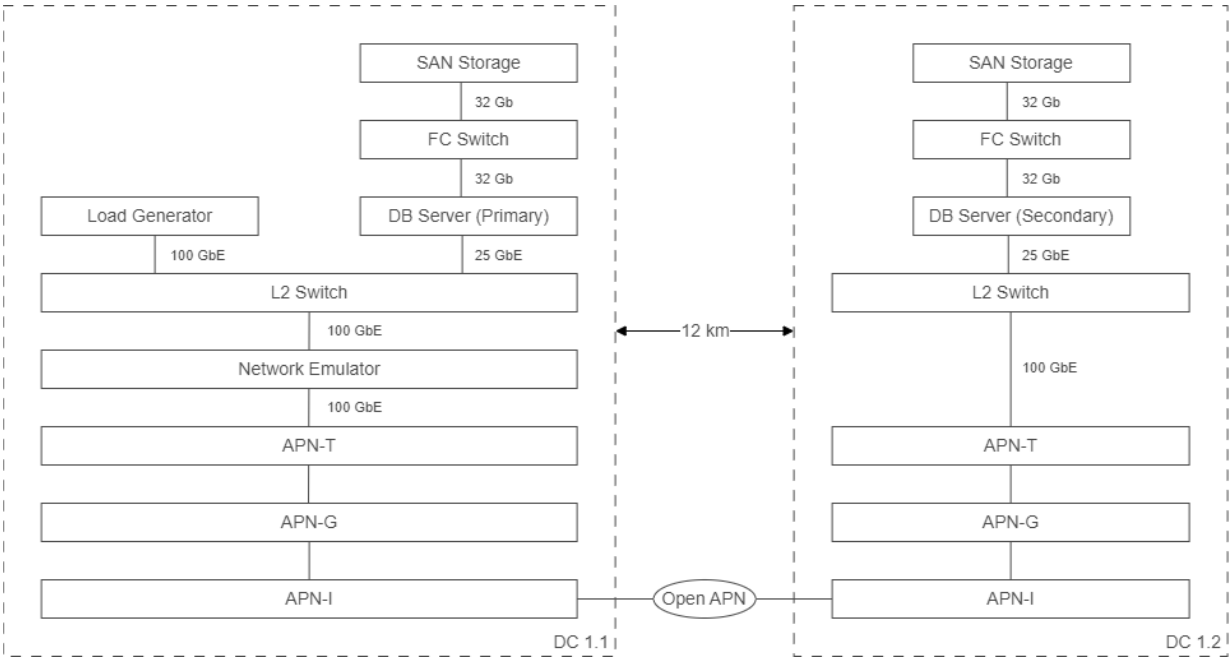

Figure 4-3: Basic Configuration

4.2.2 PoC 2: PoC Implementation

In PoC 2, the data from the primary database is replicated to the secondary database on a server located in a data center connected via Open APN. The following sections describe the implementation policies and details for each layer.

Software layer

As shown in Figure 4-4 and Table 4-4, the verification setup consists of DB Servers and a load generation tool used for testing.

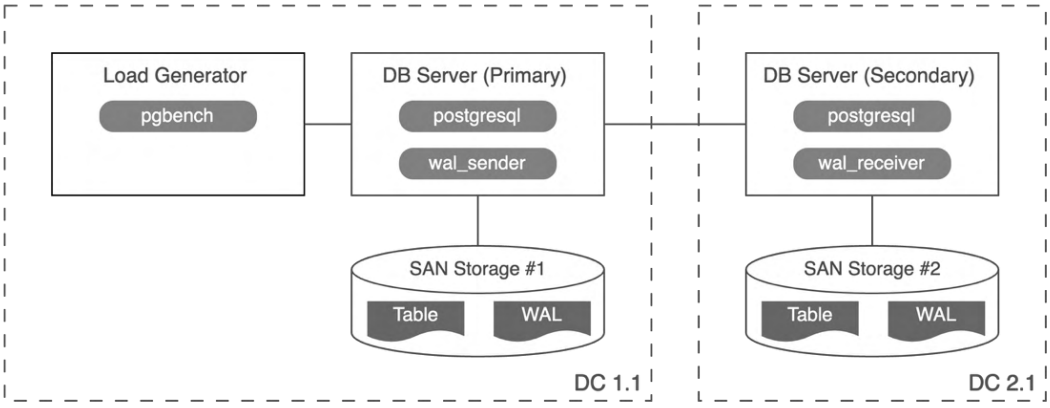

Figure 4-4: Configuration of DB replication

Table 4-4: OS and software list for PoC 2

Server	Components	Software	Version	Remarks
DB Server (Primary) DB Server (Secondary)	OS	Rocky Linux	9.3	
	DB	PostgreSQL	17.2	
		pg_wait_sampling	1.1.6	
	Device Driver	HBA Driver	14.2.566.27-1.rhel9u3	Emulex HBA Driver
Load Generator	OS	Rocky Linux	9.3	
	Language	Python	3.9.19	
	DB	pgbench	17.2	PostgreSQL benchmark tool

Network infrastructure layer

In [FSI Use Case], the minimum and maximum distances between inter-regions in Japan are defined as 250 km and 2,500 km. Accordingly, this verification was conducted over a range of 250 km to 2,500 km. However, it was not feasible to physically prepare Open APN connections covering those distances in this PoC. As shown in Figure 4-3, the only segment with an actual Open APN was the 12 km link between DC 1.1 and DC 2.1.

To simulate long-distance transmission, the time it would take for light to travel the additional distance—calculated by subtracting the actual 12 km between DC 1.1 and DC 2.1 from the desired test distance—was introduced using a network emulator. This enabled pseudo long-distance transmission equivalent to 250 km to 2,500 km.

Note 1: In the current configuration, optical amplifiers are not used because the physical optical fiber section is short.

Note 2: While the speed of light in optical fiber varies depending on various factors, a uniform value of 200,000 km/sec is used for simplicity in this PoC.

5 PoC technical report

This section describes the PoC technical Report per use case.

5.1 PoC 1: VM migration across multiple DCs

5.1.1 PoC 1: Purpose

This PoC 1 is related to intra-regional UC Scenario 1 in the use case document [FSI Use Case]. In other words, VM migration is performed from one data center to another. The migration to be performed at this time is a live migration without stopping VMs. In the following, migration refers to live migration unless otherwise specified.

In this PoC 1, VM migrations are performed per VM group, which represents a subsystem in an entire banking system. We examined the feasibility of the following Key Requirements (KRs) and features described in the Use Case document [FSI Use Case], and confirmed that they were indeed achievable:

1. The service downtime associated with executing VM migration for each VM group is less than 1 second. <KR1.1>
2. There are no lost transactions that have not been written to the database. (RPO=0)

The PoC measures the metrics specified in the RIM and PoC Reference document [FSI Use Case RIM PoC Ref]. The document includes two types of verification: Validation Testing, which is a verification to confirm the fulfillment of Key Requirements (KR) specified in the Use Case document [FSI Use Case], and Benchmark Testing, which is a verification to acquire fundamental values, etc. The better the measurements obtained through the verification, the more desirable they are. Benchmark Testing is defined as a verification to learn the fundamental values, etc., that are desirable, as the better the measurements obtained through the verification, the more desirable they are. Table 5-1 describes the metrics for this PoC 1 and the correspondence between each test.

Table 5-1: PoC 1 metrics

Symbol	Metrics	Data source (log)	VT	BT
T1	Preparation (Memory pre-copy)	VMware ESXi		Y
T2	Finalization	VMware ESXi	Y	
SVC-T1	Performance degradation time	VMware ESXi		Y
SVC-T2	Downtime	VMware ESXi	Y	
QT-LOST-HREQ	The quantity of lost HTTP requests to the Application Server	httping		
QT-SCSS-HREQ	The quantity of successful HTTP requests to the Application Server	Apache JMeter		
QT-DB-RECS	The number of DB records	PostgreSQL		

VT: Validation Testing defined in [FSI Use Case RIM PoC Ref]
BT: Benchmark Testing defined in [FSI Use Case RIM PoC Ref]

5.1.2 PoC 1: Scenario

In the environment prepared in Chapter 4, we measure performance metrics related to VM migration as described in the PoC Purpose. The measurements will be made according to scenarios focusing on two typical factors that cause performance attenuation.

Scenario 1: Performance degradation due to transmission distance

Evaluate the degree of degradation of performance metrics values such as throughput, latency, and downtime due to transmission distance. The evaluation is performed by obtaining the results of VM migrations performed within the same location and between locations connected by Open APN and comparing the results.

Scenario 2: Performance degradation due to high load

In an actual system, there is always some application running; in other words, it is normal for the system to be under load. Therefore, it is meaningful to perform VM migration when the system is under load to some extent. In this PoC, in addition to simplifying the isolation of server-side problems and the study of proposed countermeasures, an approach that places a load on the communication path is implemented to confirm the effectiveness of bandwidth control using optical wavelength path separation. Specifically, a sufficient amount of background load is generated for communication paths separated by optical wavelength path units to intentionally squeeze the bandwidth. We will then confirm that the performance attenuation that occurred during VM migration can be improved by separating optical wavelength paths for each traffic type.

Note 1: The software-defined support of Open APNs is expected to enable on-demand configuration of optical wavelength paths, enabling flexible bandwidth control in Open APNs.

Note 2: Generally, the degree of load is determined by the number of transactions to be processed, the number of simultaneous connections from clients to the system, the amount of data (number and size), and the number of parallel processes (batch processing, etc.). From a different perspective, regarding hardware resources, it is crucial to determine which resources to load: CPU, memory, disk, or network interface. On the other hand, one effective approach is to load the communication channel. Performance degradation under heavy load often complicates the problem isolation and the study of countermeasures. Still, when the communication channel is the bottleneck, problem isolation becomes relatively easy from the server's perspective.

5.1.3 PoC 1: Measurement method

This section describes the methods of measurement used in PoC 1.

Baseline

Before measuring metrics related to VM migration execution, we obtained the environment's basic performance values. The five basic performance values are network latency (RTT), network throughput (Bandwidth), network fluctuation (Jitter), and disk I/O (Bandwidth, IOPS).

Measuring latency (RTT):
The ping command is used for RTT measurement. The measurement path consists of two spatial segments: one within a DC and one between DCs. The duration is 6 hours.

```
# ping -c 10 [target_ip_addr]
```

Listing 5-1: Measuring RTT by ping command

Measuring throughput (Bandwidth):
The iperf3 command is used to measure network throughput. The measurement section is two sections, one within a DC and one between DCs. The duration is 6 hours.

```
# iperf3 -c [target_ip_addr] -P [num_of_parallel_client_streams]
```

Listing 5-2: Measuring throughput by iperf3 command

Table 5-2: Parameters in measuring throughput by iperf3 command

Parameter	Value(s)
num_of_parallel_client_streams	[1 | 2 | 4 | 8]

Measuring jitter:
The iperf3 command is used to measure network fluctuations. The measurement section consists of two sections, one within a DC and the other between DCs, and the measurement point is the server side (iperf server). The duration is 6 hours.

```
# iperf3 -u [target_ip_addr]
```

Listing 5-3: Measuring jitter by iperf3 command

Measuring disk I/O (Bandwidth, IOPS):
Measure disk I/O performance of a vSAN cluster using the fio command.

```
# fio -finename=[io_filename] \
        -rw=[io_pattern] \
        -num_jobs=4 \
        -size=10g \
        -direct=1 \
        -bs=[block_size] \
        -group_reporting \
        -name=[taskname] \
        -output=[log_filename]
```

Listing 5-4: Measuring disk io by fio command

Table 5-3: Parameter in measuring disk io by fio command

Parameter	Value(s)
io_filename	(Arbitrary value)
io_pattern	[read | write]
block_size	[8KB | 512KB | 1024KB]
taskname	(Arbitrary value)
log_filename	(Arbitrary value)

Scenario 1

Two patterns are measured: VM migration from one server to another inside DC 1.1 (Scenario 1-1, Figure 5-1) and VM migration from DC 1.1 to DC 1.2 (Scenario1-2, Figure 5-2). For VM migration, we use the vMotion functionality provided by VMware vSphere.

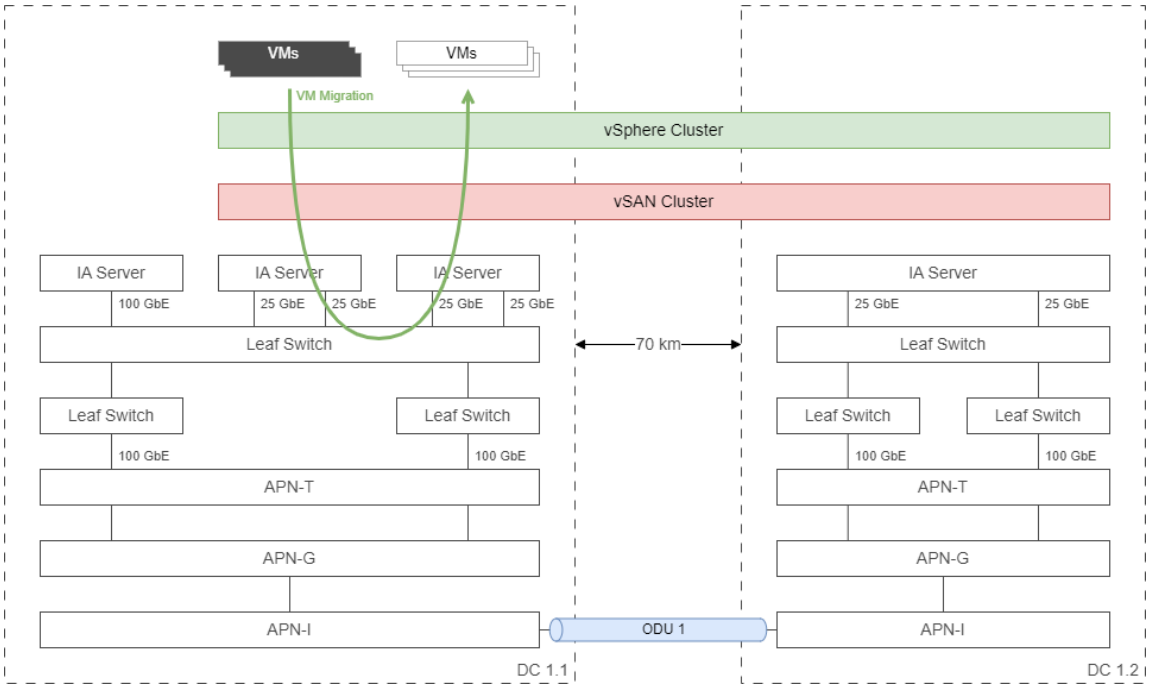

Figure 5-1: [Scenario 1-1] VM migration inside DC 1.1

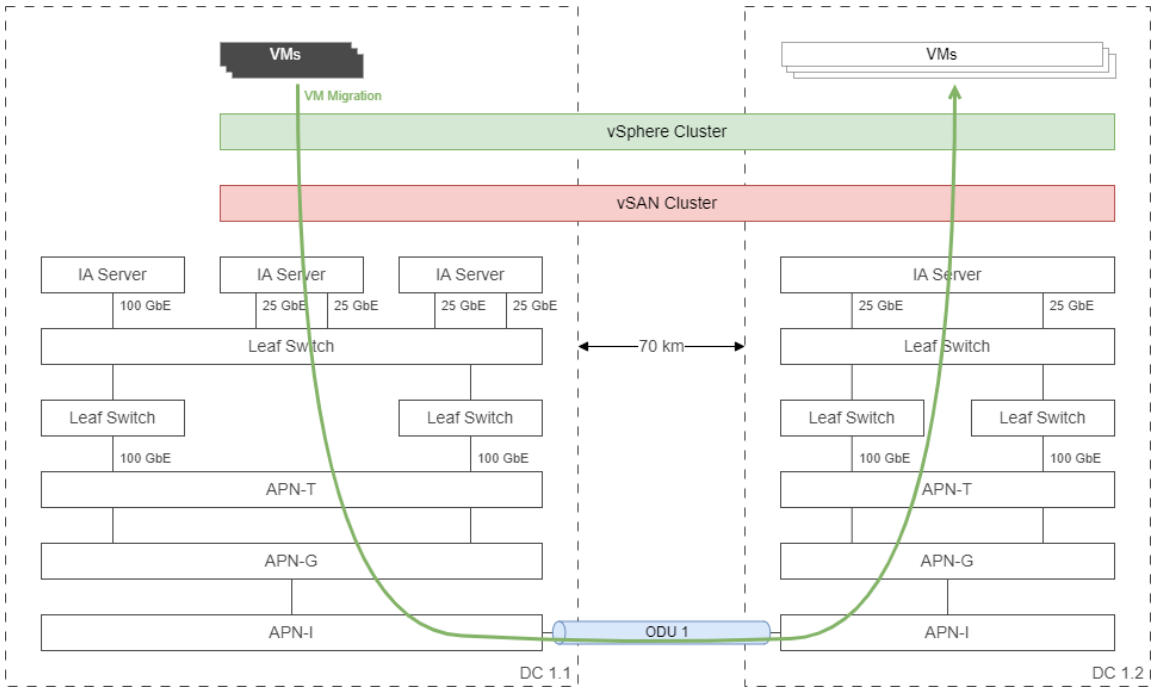

Figure 5-2: [Scenario 1-2] VM migration from DC 1.1 to DC 1.2

Scenario 2

In this scenario, a VM migration is performed from DC 1.1 to DC 1.2 with sufficient load on the vSAN cluster, using the vMotion feature provided by VMware vSphere. The ODUs that are built between DC 1.1 and DC 1.2 can be divided into two patterns: one in which all communication goes through one ODU (Scenario 2-1, Figure 5-3), and one in which two ODUs are used: one for communication with the highly loaded vSAN cluster and its communication, and the other for all vSphere communication and network load (Scenario 2-2, Figure 5-4).

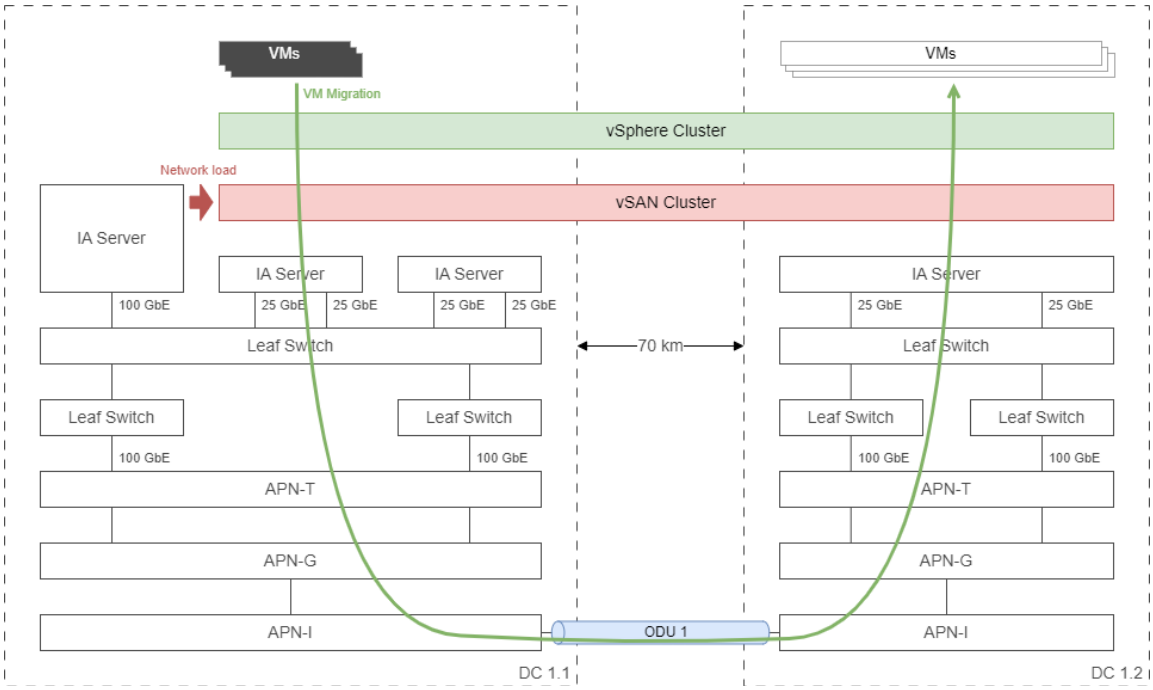

Figure 5-3: [Scenario 2-1] VM migration from DC 1.1 to DC 1.2 (w/ Network load)

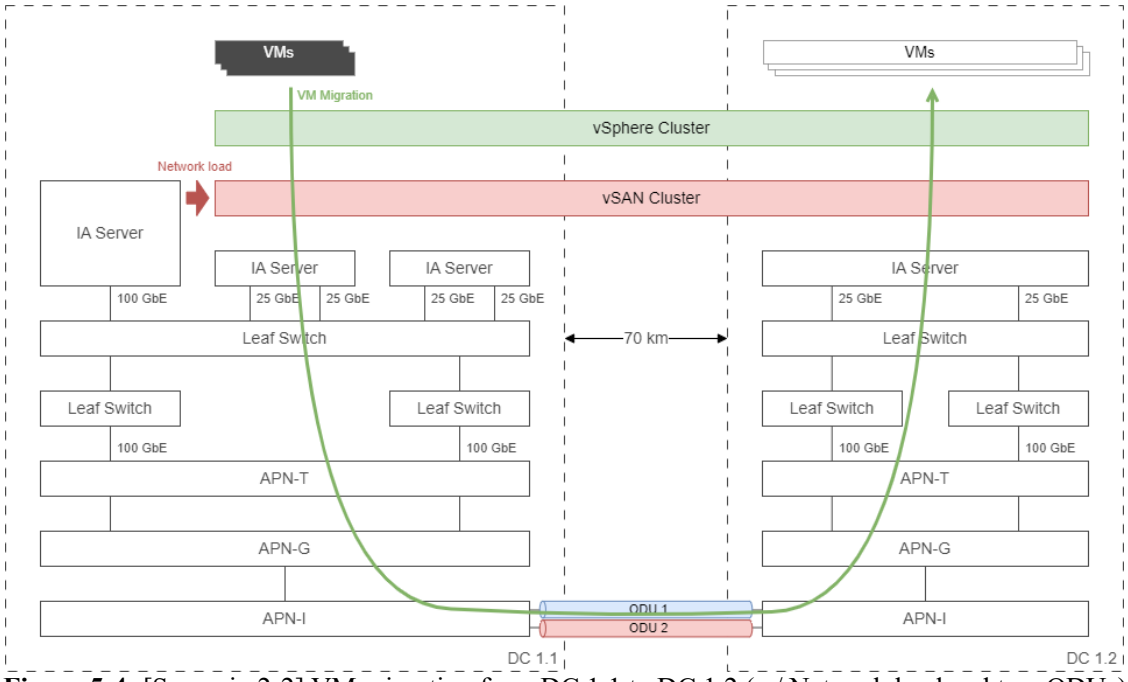

Figure 5-4: [Scenario 2-2] VM migration from DC 1.1 to DC 1.2 (w/ Network load and two ODUs)

5.1.4 PoC 1: Evaluation

Baseline

Measuring latency (RTT):
The ping command was used for RTT measurements. The measurement path consists of two spatial segments: one within a DC and one between DCs, and the duration was 6 hours.

Table 5-4: Latency (RTT) on baseline

Measurement section	Distance	Minimum [msec]	Average [msec]	Maximum [msec]
DC 1.1 to DC 1.1	<10 m (Same Rack)	0.137	0.162	0.225
DC 1.1 to DC 1.2	70 km	0.728	0.764	1.049

Measuring throughput (Bandwidth):
The iperf3 command was used to measure network throughput. The measurement section consisted of two sections, one within a DC and one between DCs, and lasted 6 hours. Since the server's NIC is 25 Gbps, we are able to measure throughput values that are almost wire-rate.

Table 5-5: Throughput (Bandwidth) on baseline

Measurement section	Parallel: 1	Parallel: 2	Parallel: 4	Parallel: 8
DC 1.1 to DC 1.1	24.7 Gbps	24.7 Gbps	24.7 Gbps	24.8 Gbps
DC 1.1 to DC 1.2	24.6 Gbps	24.7 Gbps	24.7 Gbps	24.8 Gbps

Measuring jitter:
The iperf3 command was used to measure network fluctuations. The measurement section consists of two sections, one within a DC and the other between DCs, and the measurement point was the server side (iperf server). The duration was 6 hours.

Table 5-6: Jitter on baseline

Measurement section	Distance	Minimum [msec]	Average [msec]	Maximum [msec]
DC 1.1 to DC 1.1	<10 m (Same Rack)	0.008	0.016	0.042
DC 1.1 to DC 1.2	70 km	0.006	0.021	0.115

Measuring disk I/O (Bandwidth, IOPS):
Measured disk I/O performance of a vSAN cluster using the fio command.

Table 5-7: Disk I/O (Bandwidth) on baseline

I/O pattern	Block size: 8 KB	Block size: 512 KB	Block size: 1024KB
Read	148 MB/s	1,458 MB/s	2,415 MB/s
Write	47 MB/s	1,197 MB/s	1,626 MB/s

Table 5-8: Disk I/O (IOPS) on baseline

I/O pattern	Block size: 8 KB	Block size: 512 KB	Block size: 1024KB
Read	18,900	2,915	2,414
Write	6,065	2,394	1,625

Scenario 1

We measured vMotion execution time and individual VM downtime in two patterns: vMotion from one server to another inside DC 1.1 (Scenario 1-1, Figure 5-1) and vMotion from DC 1.1 to DC 1.2 (Scenario 1-2, Figure 5-2). We measured the vMotion execution time and the downtime of each VM. The vMotion execution unit is a group of VMs that are migrated when vMotion is executed. Details are shown in Figure 5-5.

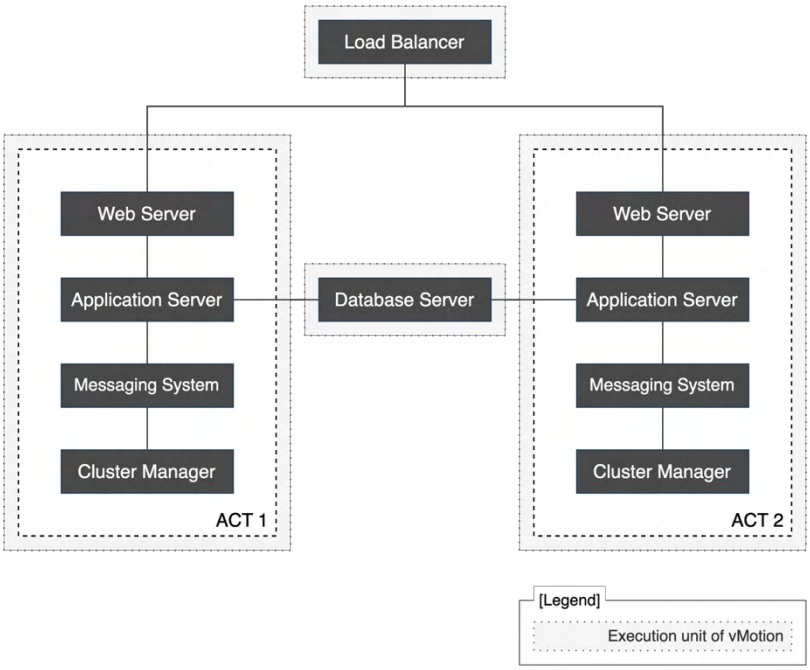

Figure 5-5: Execution unit of vMotion corresponding to the pseudo-online banking application system

Table 5-9: Result of VM migration in scenario 1

Execution unit of vMotion	VM	*T1* [msec]		*T2* [msec]		*SVC-T1* [msec]		*SVC-T2* [msec]	
		Scenario 1-1	Scenario 1-2	Scenario 1-1	Scenario 1-2	Scenario 1-1	Scenario 1-2	Scenario 1-1	Scenario 1-2
ACT 1	Web Server	7,504	10,217	119	169	≤ 24,224	≤ 29,269	≤ 528	≤ 846
	Application Server	7,992	9,727	170	340				
	Messaging System	4,613	4,953	134	182				
	Cluster Manager	4,115	4,372	105	155				
ACT 2	Web Server	6,730	8,485	131	176	≤ 23,848	≤ 29,267	≤ 574	≤ 864
	Application Server	6,965	10,056	181	346				
	Messaging System	5,023	5,341	151	180				
	Cluster Manager	5,130	5,385	111	162				
Database Server	Database Server	25,988	11,755	118	170	25,988	11,755	118	170
Load Balancer	Load Balancer	5,244	8,206	109	181	5,244	8,206	109	181

Note: Metris defined in Table 5-1.

Table 5-9 shows the measured results of vMotion processing time and downtime for the evaluations of Scenario 1-1 and 1-2. In comparison to Scenario 1-1, many VMs in Scenario 1-2 had longer vMotion processing times and experienced more downtime. VMs with higher workloads—such as web servers, application servers, databases, and load balancers—tended to have increased times. As shown in the Figure 5-6, an accurate evaluation of service downtime requires subtracting overlapping periods from the sum of IT downtimes across all VMs. In this evaluation, no adjustment for overlapping durations was applied; therefore, service downtime was conservatively assessed as being less than or equal to the total sum of individual IT downtimes. This is what the "≤" in Table 5-9 represents.

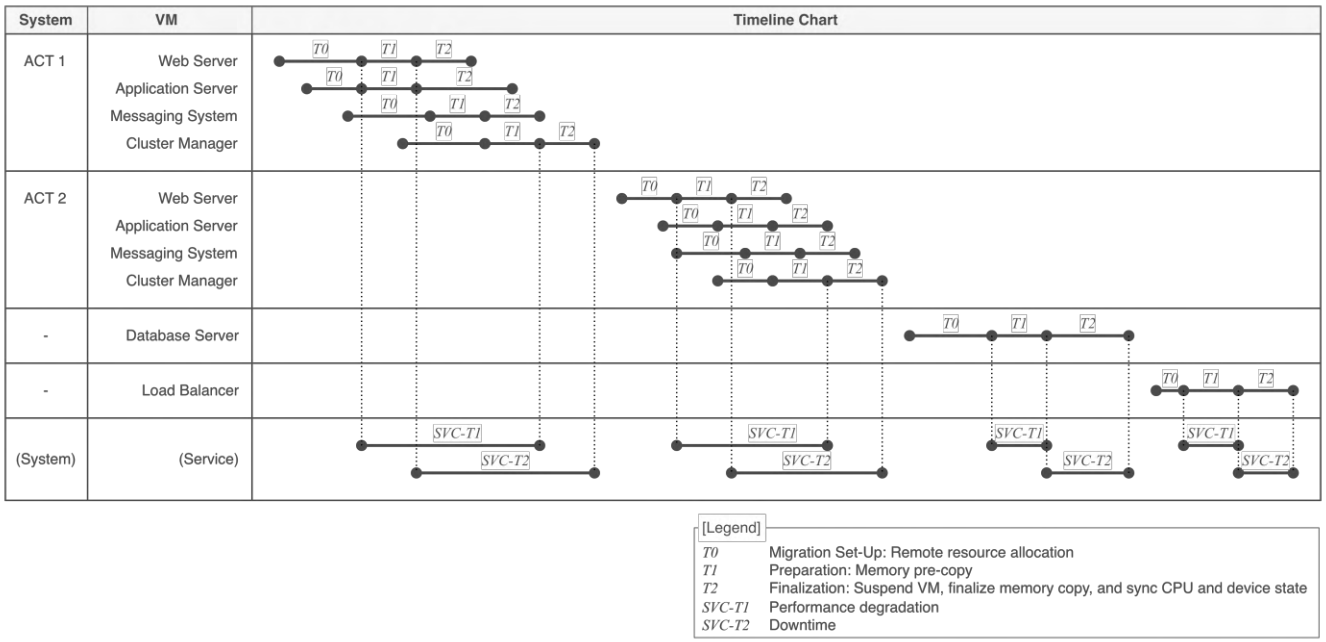

Figure 5-6: Sequence diagram of VM migration

Table 5-10: Result of table data consistency verification in scenario 1

Component	Metrics		Result	
			Scenario 1-1	Scenario 1-2
Load Generator	The number of HTTP requests	All	261,923	177,475
		Succeeded	261,923	177,472
		Failed	0	3
	HTTP requests/sec		1,463	1,133
Database Server	The number of inserted records into *acc_account_activity* table	Deposit	261,923	177,472
		Withdraw	261,923	177,472
	The number of inserted records into acc_money_transfer table	Transfer	261,923	177,472

Table 5-10 shows the comparison of the number of successful HTTP requests by Load Generator (Apache JMeter) and the number of DB records. The client simply waits for responses from the server. The objective of this PoC is to verify the feasibility of [FSI Use Case RIM PoC Ref]. To this end, we validated the key requirements and benchmarked several performance indicators within a minimal VM-based system configuration. Note that the total number of HTTP requests differs between Scenarios 1-1 and 1-2, but this is due to the difference in measurement time. Although some HTTP requests failed in Scenario 1-2, this is not a problem because the number of records inserted in the database matches the number of successful HTTP requests. As a result, there are no lost transactions that have not been written to the database.

Scenario 2

vMotion was performed from DC 1.1 to DC 1.2 with sufficient load on the network interface of the vSAN cluster. At this time, we measured two patterns in the ODUs built between DC 1.1 and DC 1.2: when all communication goes through one ODU (Scenario 2-1, Figure 5-3), and when two ODUs are used: one for the highly loaded vSAN cluster and its communication, and the other for other communication (Scenario 2-2, Figure 5-4). The vMotion execution unit in Scenario 2 is the same as in Scenario 1.

Table 5-11: Result of VM migration in scenario 2

Execution unit of vMotion	VM	*T1* [msec]		*T2* [msec]		*SVC-T1* [msec]		*SVC-T2* [msec]	
		Scenario 2-1	Scenario 2-2	Scenario 2-1	Scenario 2-2	Scenario 2-1	Scenario 2-2	Scenario 2-1	Scenario 2-2
ACT 1	Web Server	14,980	9,998	261	195	≤ 37,903	≤ 29,787	≤ 1,337	≤ 941
	Application Server	11,651	10,392	595	350				
	Messaging System	6,023	4,959	235	187				
	Cluster Manager	5,249	4,438	246	209				
ACT 2	Web Server	10,166	8,417	246	223	≤ 35,258	≤ 25,911	≤ 1,265	≤ 988
	Application Server	12,025	9,061	549	389				
	Messaging System	7,061	4,233	243	213				
	Cluster Manager	6,106	4,200	227	163				
Database Server	Database Server	30,740	11,770	330	214	30,740	11,770	330	214
Load Balancer	Load Balancer	29,414	8,663	366	218	29,414	8,663	366	218

Note: Metrics defined in Table 5-1.

Table 5-11 shows the measured vMotion processing times and downtimes for Scenarios 2-1 and 2-2. Compared to Scenario 2-1, all VMs in Scenario 2-2 showed shorter vMotion processing times and downtimes. As shown in the Figure 5-6, an accurate evaluation of service downtime requires subtracting overlapping periods from the sum of IT downtimes across all VMs. In this evaluation, no adjustment for overlapping durations was applied; therefore, service downtime was conservatively assessed as being less than or equal to the total sum of individual IT downtimes. This is what the "≤" in Table 5-11 represents.

Table 5-12: Result of table data consistency verification in scenario 2

Component	Metrics		Result	
			Scenario 2-1	Scenario 2-2
Load Generator	The number of HTTP requests	All	230,389	214,679
		Succeeded	230,388	214,676
		Failed	1	3
	HTTP requests/sec		1,097	1,156
Database Server	The number of inserted records into *acc_account_activity* table	Deposit	230,388	214,676
		Withdraw	230,388	214,676
	The number of inserted records into acc_money_transfer table	Transfer	230,388	214,676

Table 5-12 compares the number of successful HTTP requests and the number of DB records by Load Generator (Apache JMeter). As in Scenario 1, the number of lost transactions was 0.

Summary

The testing results are presented in Table 5-9 and Table 5-11, summarizing the observed downtimes under various scenarios. A critical requirement for the system is that downtime should remain below 1 second, a condition proven feasible within most scenarios tested.

In three of the four scenarios tested, the total downtime remained under 1 second, effectively meeting the requirement. However, in Scenario 2-1, where intensive background network traffic was generated, reaching up to 100 Gbps and imposing additional load on both the vSAN cluster and the Open APN system, the downtime during the migrations of ACT 1 and ACT 2 exceeded 1 second. This scenario represents a stress test designed to simulate an undesired situation. Thus, the results of Scenario 2-1 are not problematic and can be considered anticipated outcomes under such extreme conditions.

These findings highlight the system's capability to achieve low downtimes under normal operating conditions while demonstrating predictable behavior in high-stress situations.

Careful attention is required when measuring the downtime associated with system-level migrations. While measuring *T2*: Finalization of each VM is relatively straightforward, evaluating service downtime introduces complexities and involves consideration of various factors, including the actual impact on users. *SVC-T2*: Downtime is influenced by the frequency and volume of client requests, which vary across different systems. This type of measurement often demands specialized techniques tailored to specific circumstances. In this PoC, the composite value of each IT downtime shown in the figure was used, and the actual IT downtime was evaluated as being less than or equal to this composite value. IT downtime is proposed to be represented on a timeline chart, while service downtime is defined as the cumulative downtime, excluding overlapping periods across multiple VMs.

5.2 PoC 2: Long-distance DB sync/async replication

5.2.1 PoC 2: Purpose

This PoC 2 is related to the inter-region use case defined in [FSI Use Case], and conducted following [FSI Use Case RIM PoC Ref], which has two types of testing below.

- Validation Testing: Perform asynchronous replication processing and measure "Replication Lag" and "Grey Data".
- Benchmark Testing: Perform synchronous replication processing and asynchronous replication processing, respectively, and measure "DB Performance, in Transactions / Sec (for Tier 1)".

As defined in [FSI Use Case RIM PoC Ref], these two testing types require the measurement of several metrics. Metrics include those obtained through DB synchronous replication, DB asynchronous replication, and other methods. We measured some of the metrics regarding DB synchronous replication and asynchronous replication in our PoC. (Table 5-13).

Table 5-13: PoC 2 test items and metrics

Test types	Test items		Data source (log)	Replication types	
	ID	Metrics		Sync.	Async.
Validation Testing	V-1	Replication lag (=RPO)	PostgreSQL*1		Y
	V-2	Number of unreplicated data (Grey Data)	PostgreSQL*2 and pgbench		Y
Benchmark Testing	B-1	*TATtrxn* - Latency from an application perspective	pgbench	Y	Y
	B-2	*Ptrxn* - Throughput from an application perspective	pgbench	Y	Y
	B-3	*TATrsync* - Latency of synchronous replication operations	PostgreSQL*1	Y	
	B-4	*TATrasync* - Latency of asynchronous replication operations	PostgreSQL*3		Y
	B-5	*Prasync* - Throughput of asynchronous replication operations	PostgreSQL*4		Y

*1: Column “flush_lag” of view “pg_stat_replication”.
*2: Column “flush_lsn” and “send_lsn” of view “pg_stat_replication” and column “wal_bytes” of view “pg_stats_statements”.
*3: Column “flush_lag” and “write_lag” of view “pg_stat_replication”.
*4: Column “flush_lsn” of view “pg_stat_replication”.

5.2.2 PoC 2: Scenario

The test scenarios in PoC 2 correspond to the DB replication types as follows.

- Scenario 1: Synchronous replication
- Scenario 2: Asynchronous replication

Note: It was more efficient to execute testing following categorization in each type of DB replication than for each test type.

Scenario 1: Synchronous replication

In PostgreSQL synchronous replication, to ensure that committed data is not lost even in a disaster (RPO=0), the primary database waits for WAL transfer completion to the standby database. However, this adversely affects "performance as seen from the application," making WAL transfer processing (especially data transmission delay between data centers, as shown in the red area Figure 5-7) significantly impact overall performance.

Generally, in synchronous replication, the shorter the distance between DBMSs, the shorter the synchronization wait time can be expected. Therefore, using Open APN between DC 1.1 and DC 2.1 is expected to make the wait time approach the light transmission delay as closely as possible.

Scenario 2: Asynchronous replication

In PostgreSQL asynchronous replication, the primary DB does not wait for the completion of WAL transfer to the standby database, so the "performance from the application's point of view" does not depend on the network communication performance between DCs.

When a failure occurs on the primary database side, it is necessary to recover the data from a location other than the database for data that could not be replicated to the standby database. By using a line with lower network latency, the replication lag (i.e., the time from when data is written to the primary until it is written to the secondary) is shortened. As a result, the number of transactions in the middle of replication is expected to decrease. Therefore, by using Open APN between DC 1.1 and DC 2.1, it is expected to reduce the amount of data to be recovered (=RTO shortening).

5.2.3 PoC 2: Measurement method

Scenario 1: Synchronous replication

The evaluation method involves generating a constant transaction load on PostgreSQL using pgbench. During the evaluation, the distance of the Open APN is varied to observe how this affects the measured metrics. The impact of changing load conditions—such as the number of concurrent connections—is evaluated to understand how different situations affect performance results. The relationship between the sequence diagrams for synchronous replication processing and the items measured in this verification is shown in Figure 5-7.

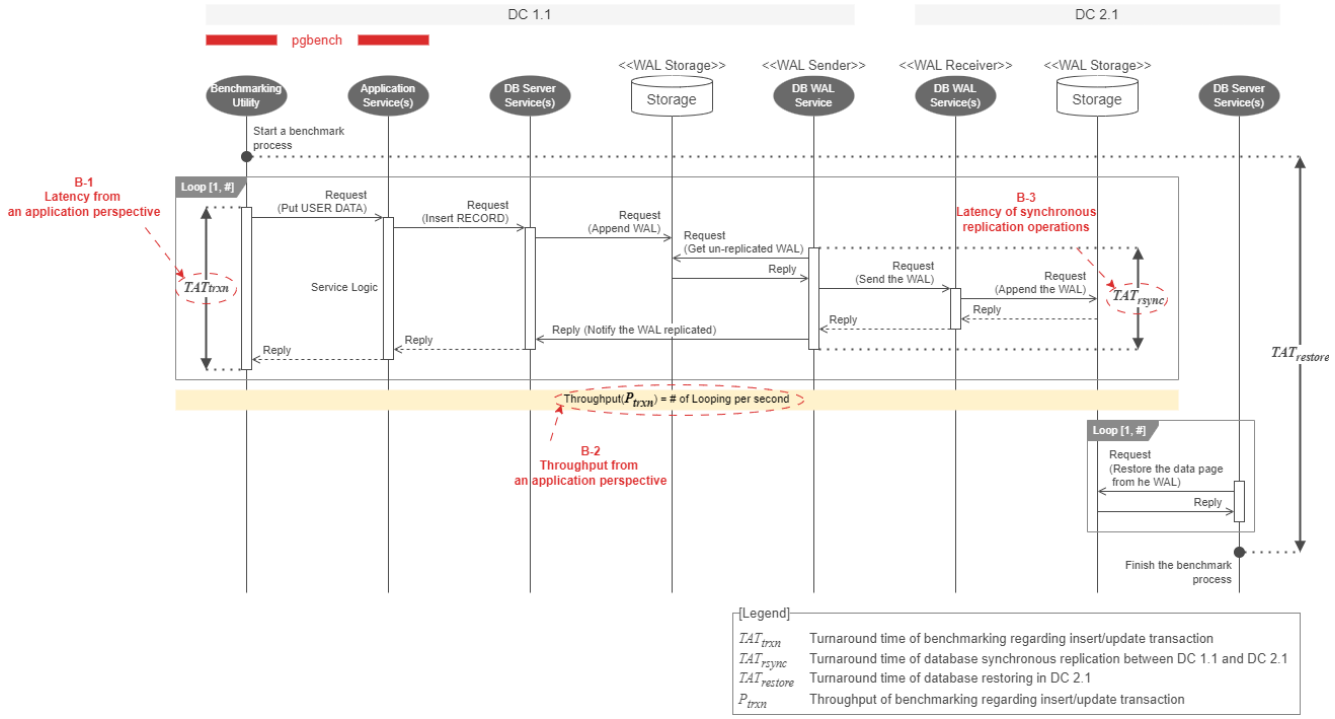

Figure 5-7: Sequence diagram of DB synchronous replication for scenario #1 of Use Case 2 [FSI Use Case RIM PoC Ref]

Scenario 2: Asynchronous replication

As in Scenario 1, the evaluation method involves generating a constant transaction load on PostgreSQL using pgbench. During the evaluation, the distance of the Open APN is varied to observe how this affects the measured metrics. The impact of changing load conditions—such as the number of concurrent connections—is evaluated to understand how different situations affect performance results. The relationship between the sequence diagram used in the asynchronous replication process and the items measured in this verification is shown in Figure 5-8.

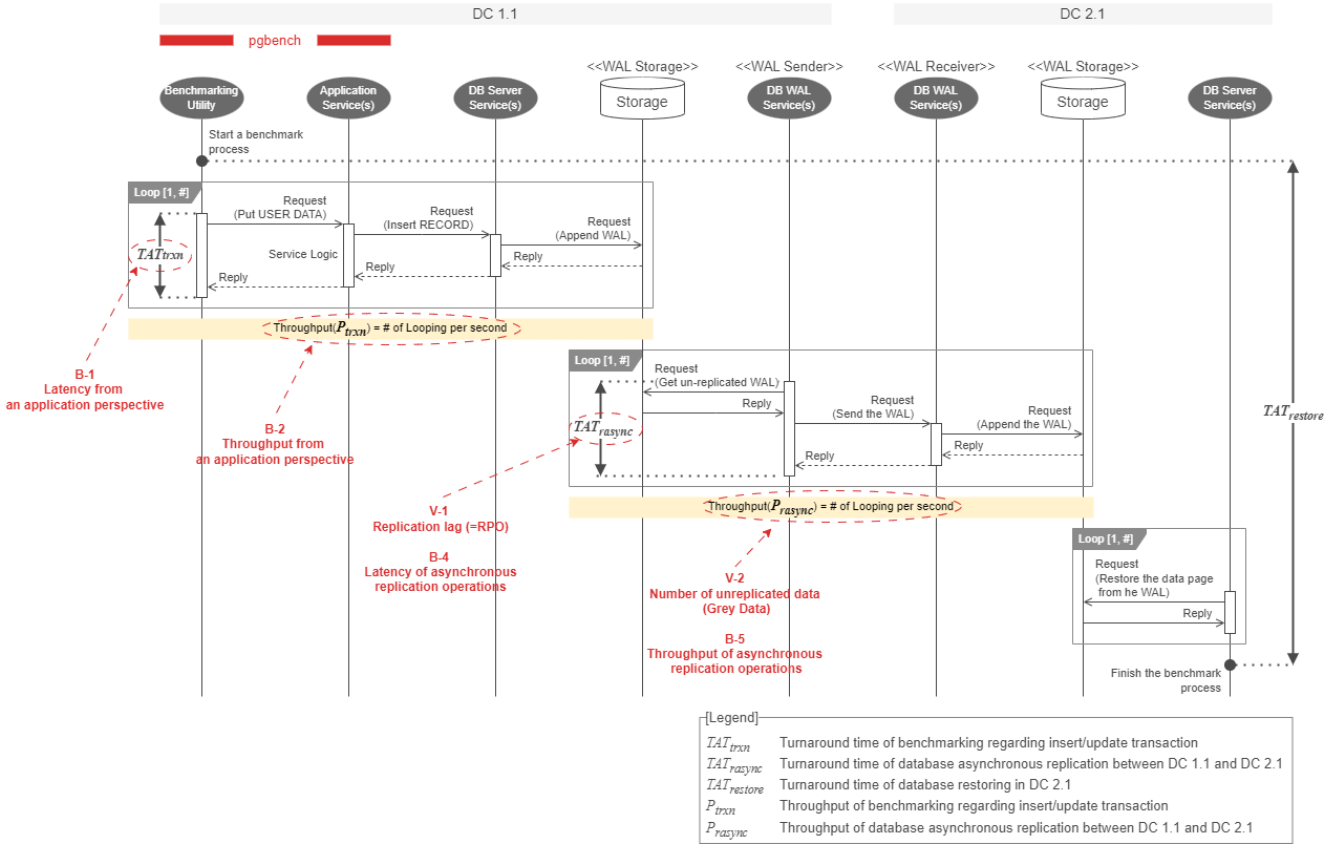

Figure 5-8: Sequence diagram of DB asynchronous replication for scenario #1 of Use Case 2 [FSI Use Case RIM PoC Ref]

5.2.4 PoC 2: Evaluation

Scenario 1: Synchronous replication

[Test ID: B-1] Measuring latency from an application perspective (*TATtrxn*)
In this verification, it can be regarded as latency from the perspective of Application Services. Figure 5-9 shows the measurement results for latency as observed from the application. It was confirmed that the average latency from the application also increased proportionally to the distance between the DCs. Up to multiplicity 300, the delay setting time accounted for most of the latency, with network transmission delay being the bottleneck. It was confirmed that the latency for multiplicities 600 and 900 increased significantly compared to multiplicities of 300 or less.

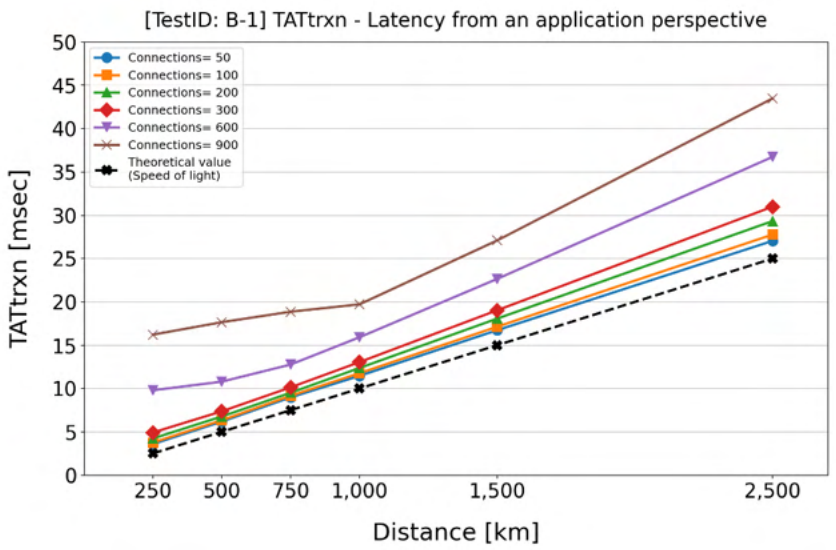

Figure 5-9: [Test ID: B-1] *TATtrxn* - Latency from an application perspective

[Test ID: B-2] Measuring throughput from an application perspective (*Ptrxn*)
We measured the value of Transaction Per Second (TPS), which is the benchmark result of pgbench. Figure 5-10 shows application throughput measurements. With an equal number of concurrent connections, average throughput decreases as the distance between DCs increases. Increasing the multiplicity confirmed that the decrease in throughput could be suppressed even when the distance between DCs increases.

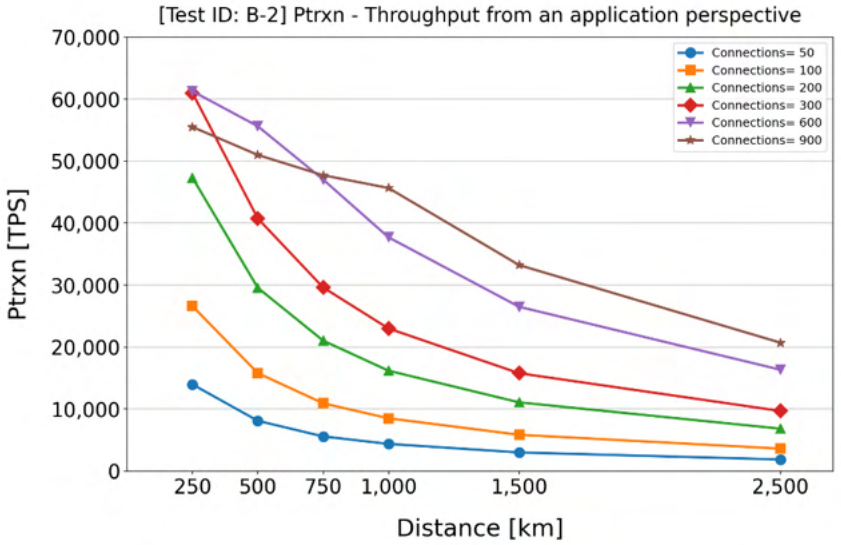

Figure 5-10: [Test ID: B-2] *Ptrxn* - Throughput from an application perspective

[Test ID: B-3] Measuring latency of synchronous replication operations (*Ptrxn*)
We measured the average values of flush_lag and write_lag from pg_stat_replication, which are operating statistics of PostgreSQL, are defined as replication lag, and this information is obtained at regular intervals to calculate the average value. Figure 5-11 presents the latency measurements for synchronous replication. The average latency of synchronous replication was nearly equal to the delay setting time (round trip), and it was confirmed that the transmission delay was the main factor.

The network transmission delay is the bottleneck, and the average latency of synchronous replication increases proportionally with the distance between the DCs.

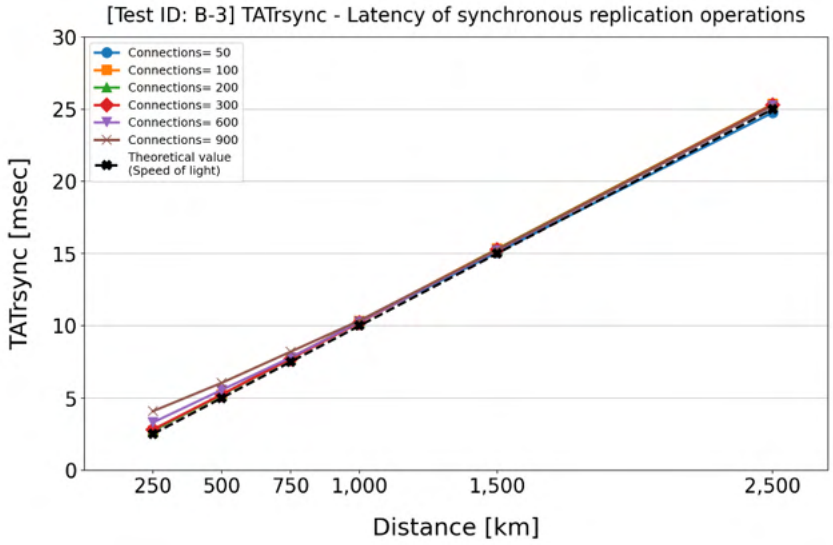

Figure 5-11: [Test ID: B-3] *TATrsync* - Latency of synchronous replication operations

Scenario 2: Asynchronous Replication

[Test ID: V-1] Measuring replication lag (=RPO)
We measured the average replication lag periodically retrieved from PostgreSQL. Figure 5-12 shows the measurement results for Lag in Asynchronous Replication. The average lag was almost the same as the delay setting time (round trip time), and it was confirmed that the transmission delay was the dominant term.

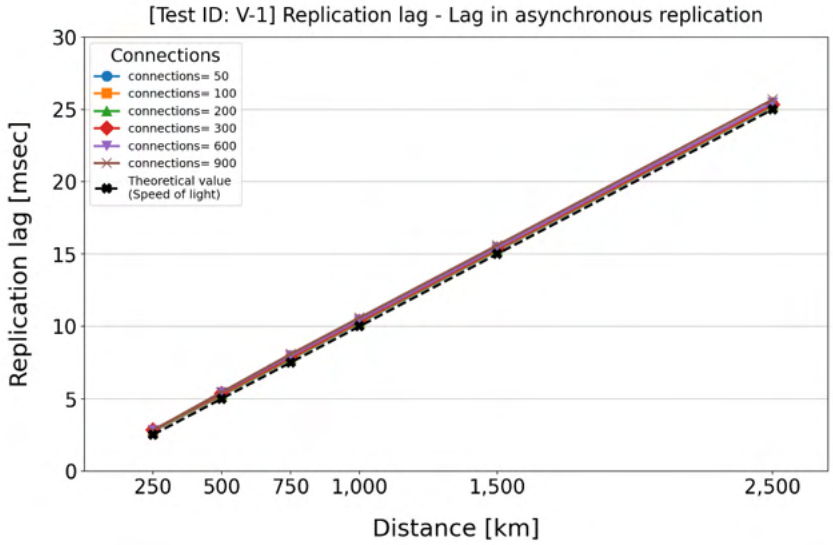

Figure 5-12: [Test ID: V-1] Replication lag (=RPO)

[Test ID: V-2] Measuring number of unreplicated data (Grey Data)
We measured the average number of unreplicated transactions, periodically collected from PostgreSQL, which indicates the amount of data that would need to be recovered from sources other than the database in case of a failure. Figure 5-13 shows the measurement results for the number of transactions not replicated in Asynchronous Replication. It was confirmed that the number of transactions not replicated at that moment (the amount of grey data) increased in proportion to the distance between data centers. In addition, it was confirmed that, up to 300 simultaneous connections, the number of unreplicated transactions also increased as the number of simultaneous connections increased.

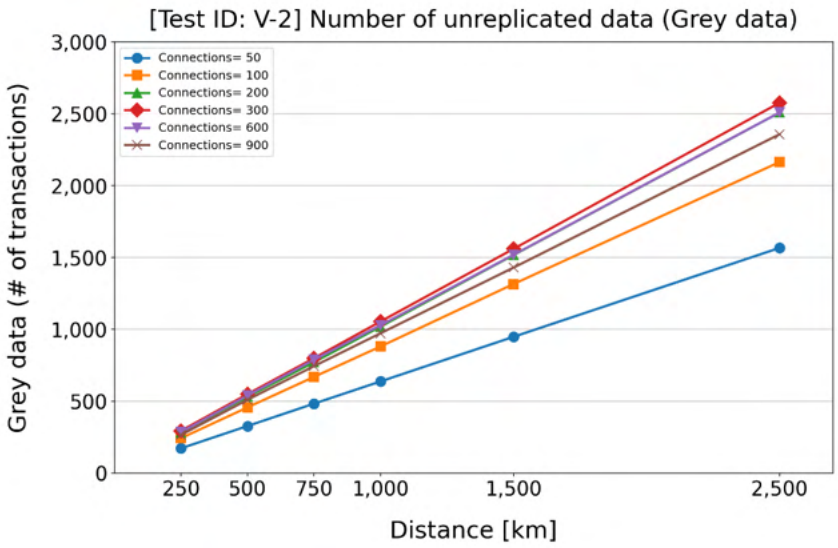

Figure 5-13: [Test ID: V-2] Number of unreplicated data (Grey Data)

[Test ID: B-1] Measuring latency from an application perspective (*TATtrxn*)
We measured the value of the latency average, which is the benchmark result of pgbench. In this verification, it can be regarded as latency from the perspective of application service(s).
Figure 5-14 shows the application latency results. Average latency remained nearly constant regardless of the distance between DCs. It was confirmed that the average latency did not decrease as the distance between DCs, and that it stayed almost constant. It was also confirmed that the average latency for multiplicities of 600 and 900 increased compared to multiplicities of 300 or less.

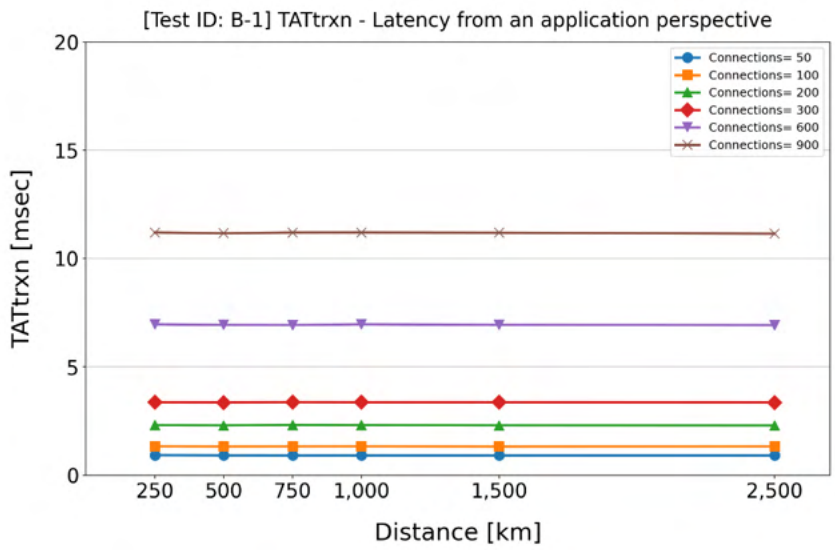

Figure 5-14: [Test ID: B-1] *TATtrxn* - Latency from an application perspective

[Test ID: B-2] Measuring throughput from an application perspective (*Ptrxn*)
We measured the value of TPS, which is the benchmark result of pgbench. Figure 5-15 shows the measurement results for Throughput from the application's perspective. It is confirmed that the average throughput remains almost constant while the distance between DCs increases. It was also confirmed that the throughput of multiplicity 600 and 900 decreased compared to multiplicity 300 or less.

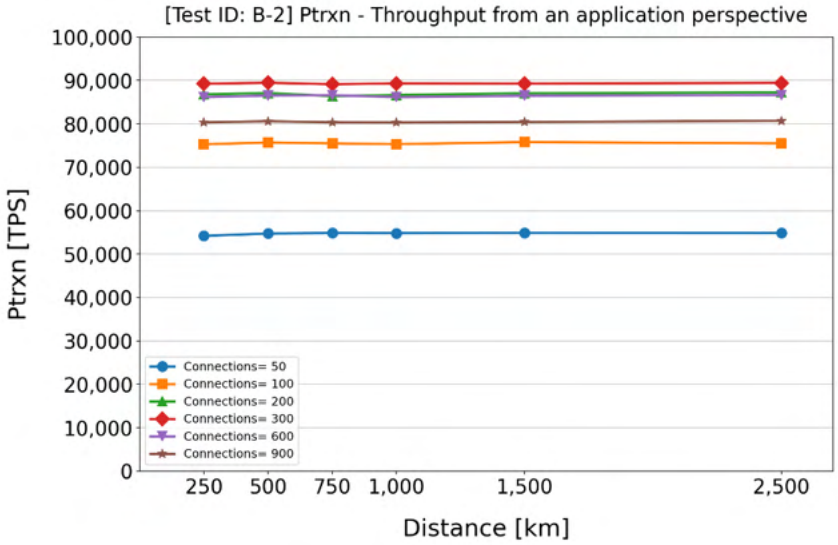

Figure 5-15: [Test ID: B-2] *Ptrxn* - Throughput from the perspective of the application

[Test ID: B-4] Measuring latency of asynchronous replication processing (*TATrasync*)
The evaluation uses the same metric as “lag in asynchronous replication (V-1),” which is the average number of replication WAL transfers per second, periodically retrieved from PostgreSQL. Figure 5-16 presents the measurement results for the latency of asynchronous replication processing. The average latency of asynchronous replication was nearly the same as the round-trip delay setting, confirming that transmission delay is the main factor. However, when the system is overloaded (with 600 or 900 simultaneous connections) and the distance to the data center exceeds 1500 km, a temporary spike in average replication latency is observed. Therefore, when

calculating average latency, the point of rapid increase is considered an outlier and is excluded from the calculation

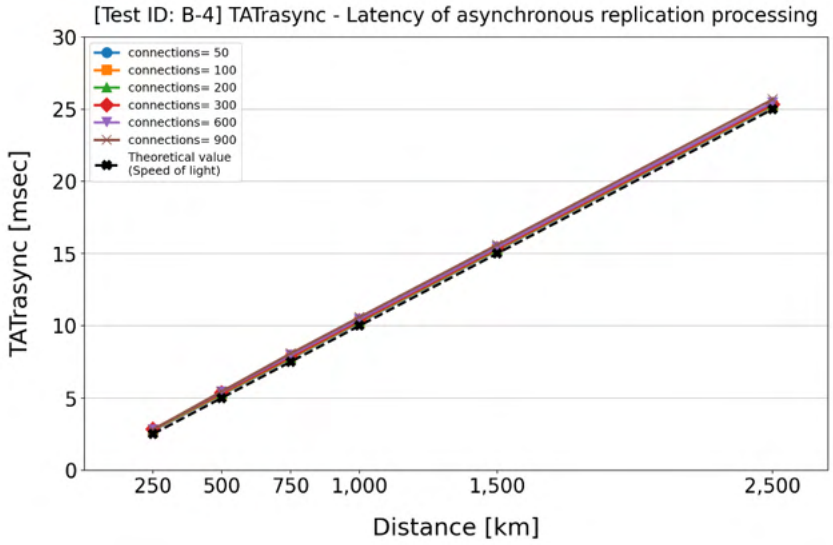

Figure 5-16: [Test ID: B-4] *TATrasync* - Latency of asynchronous replication processing

[Test ID: B-5] Measuring throughput of asynchronous replication processing (*Prasync*)
We measured the average replication WAL transfers/sec retrieved from PostgreSQL. Figure 5-17presents the throughput results for asynchronous replication processing. Regardless of the distance between DCs, the average throughput of asynchronous replication (WAL transfer volume/sec) remained nearly constant and stable.

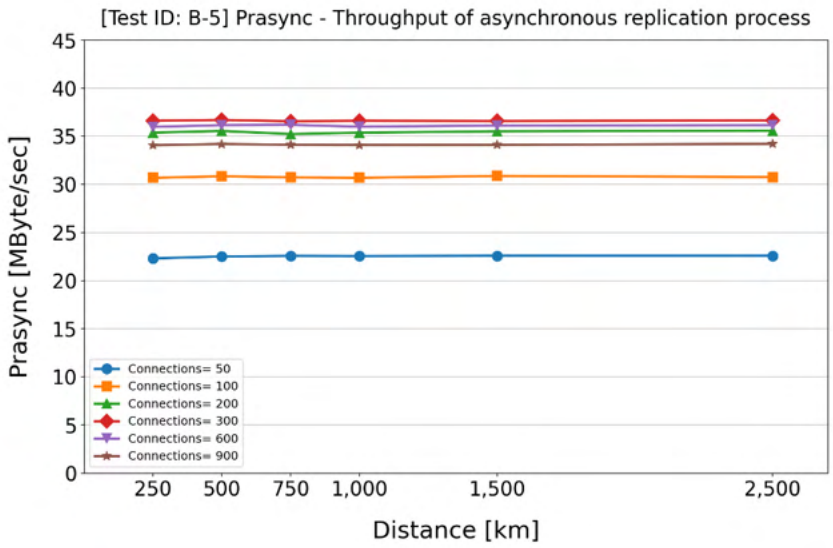

Figure 5-17: [Test ID: B-5] *Prasync* - Throughput of asynchronous replication processing

Summary

The results of synchronous replication demonstrate the following characteristics:

- As expected, application performance (throughput and latency) was observed to decline as the distance between data centers increased. This is due to the primary database waiting for the transfer of Write-Ahead Logging (WAL) to the standby DB during transaction commit, which dominates processing time.
- It was confirmed that increasing the number of simultaneous connections to enhance parallel processing could reduce the decline in throughput, even as DC distances grew.

Traditionally, synchronous database replication has been considered unsuitable for long-distance environments exceeding 100 km and regarded as a typical anti-pattern. However, the results of this PoC revealed that under a wide range of conditions—such as within 500 km and with 300 or fewer concurrent connections—synchronous replication, as observed from the application side, can be completed in under 10 milliseconds. This indicates the potential to extend the feasible distance for synchronous replication—possibly even overturning the conventional understanding of this anti-pattern.

The results of asynchronous replication demonstrate the following features:

- As expected, application performance (throughput and latency) stayed consistent regardless of the DC distance. There are no replication wait times, which usually slow down processing in synchronous replication.
- It was observed that increasing the DC distance proportionally results in a larger volume of "grey data" needed for external DB recovery during disaster scenarios.
    - In long-distance setups (e.g., 2,500 km), the grey data volume remains limited to about 2,500 transactions, and the data can be processed efficiently at 90,000 transactions per second (TPS), enabling the complete DB storage in less than a second.

A comparison of the results between synchronous and asynchronous replication confirmed that asynchronous replication becomes increasingly advantageous in terms of performance as the distance between data centers grows. However, it was also confirmed that the occurrence of grey data is inevitable and tends to increase with greater distance.

The measurement results are evaluated in terms of whether they meet the Key Requirements(KR) of the [FSI Use Case].

- KR2.1: With synchronous replication, the primary DB waits for the write to the standby DB, resulting in RPO = 0. On the other hand, with asynchronous replication, it was confirmed that RPO depends on the transmission distance.
- KR2.2: With synchronous replication, since the primary DB waits for the write to the standby DB, no data remains unreplicated in the standby DB. In contrast, with asynchronous replication, unreplicated transactions (grey data) occur in the standby DB, but even over a long distance (2,500 km), the number was about 2,500 transactions. As the recoverable data can be clearly identified and assuming processing at 90,000 TPS (see Figure 5-15), the storage process to the DB itself is considered to be executable in less than one second.

6 Contribution to the IOWN Global Forum

- RIM-TF
  - The PoC supports the Reference Implementation Model for distributed data center systems for financial institutions.

7 Conclusion

This PoC was conducted to validate that the key requirements (KR) defined in the [FSI Use Case] and [FSI Use Case RIM PoC Ref] are achievable within a simulated financial system environment.

PoC 1: VM Migration

Across all test scenarios, VM migration completed with downtime of under one second, satisfying the KR criteria. In client–server systems, end users typically do not notice downtime shorter than one second and do not consider it a system failure. This result demonstrates the potential for financial system operators to relocate and reconfigure server-side application components flexibly without compromising the user experience. Consequently, the findings suggest the practical viability of dynamic resource management and more efficient utilization of ICT resources within a single region.

PoC 2: Database Replication

During the validation of synchronous replication, RPO = 0—as required by the KRs—was achieved between data centers in different regions, ensuring data consistency. These results show that synchronous database replication enables real-time synchronization across multiple regional sites. This breakthrough points toward a transformation of traditional system architectures designed to ensure data redundancy, marking an initial step in redefining next-generation financial systems.

This PoC confirmed that the KRs defined in the [FSI Use Case] and [FSI Use Case RIM PoC Ref] are achievable within a simulated financial system environment. The findings suggest that it is possible to alleviate traditional geographic and operational constraints, leading to enhanced system availability, broader Business Continuity Planning and Disaster Recovery design options, and greater flexibility in resource utilization for financial systems.

References

FSI Use Case	Services Infrastructure for Financial Industry Use Case https://iowngf.org/services-infrastructure-for-financial-industry-use-case/
FSI Use Case RIM PoC Ref	Reference Implementation Model and Proof-of-Concept Reference of Services Infrastructure for Financial Industry Use Case https://iowngf.org/reference-implementation-modeland-proof-of-concept-referenceof-services-infrastructure-forfinancial-industry-use-case/

Appendix A. Suggestions for future improvements: FC-SAN over Open APN

A.1 Introduction of FC-SAN over Open APN

This section presents the results of testing involving the geographic extension of a Fibre Channel-based Storage Area Network (FC-SAN) over Open APN, hereafter called FC-SAN over APN. Generally, SANs have been designed with the assumption that all components are located within the same data center, and FC-SANs, in particular, have faced significant technical challenges to geographic distribution because Fibre Channel's protocol was not originally designed for long-distance communication.

In this PoC, we successfully interconnected data centers separated by 70 to 120 kilometers via Open APN and verified that storage read/write operations could be performed with stable performance. These results not only demonstrate the feasibility of long-distance Fibre Channel communication but also highlight the potential for geographically decoupled SAN architectures. Although the current [FSI Use Case] and [FSI Use Case RIM PoC Ref] do not yet include SAN scenarios, this report is provided as a reference with the expectation that support for FC over Open APN will be considered in future IOWN Global Forum initiatives.

A.2 Motivation of FC-SAN over Open APN

Mission-critical ICT systems require high reliability, such as data deliverability assurance and lossless data to deliver financial services correctly. FC-SAN is a storage network architecture deployed inside the DC that provides high throughput and low latency. Open APN and FC-SAN have similar network characteristics, so their interoperability will inherit their advantages and will lead to a wider range of FC-SAN usage.

As shown in Figure A-1, SAN (Storage Area Network) is typically a network built within a data center since FC is not intended for long-distance use. Therefore, with the current FC-SAN setup, it is difficult to distribute the DB server and storage geographically. However, if long-distance Fibre Channel support is enabled, it would become possible to separate the DB server and storage over long distances. This geographic flexibility of SAN storage could offer the following benefits:

1. Enables flexible system design with secure data, supporting both cloud and local applications.
2. Supports virtualization of SQL and DB functions for better scalability.
3. Ensures guaranteed, lossless long-distance data delivery.

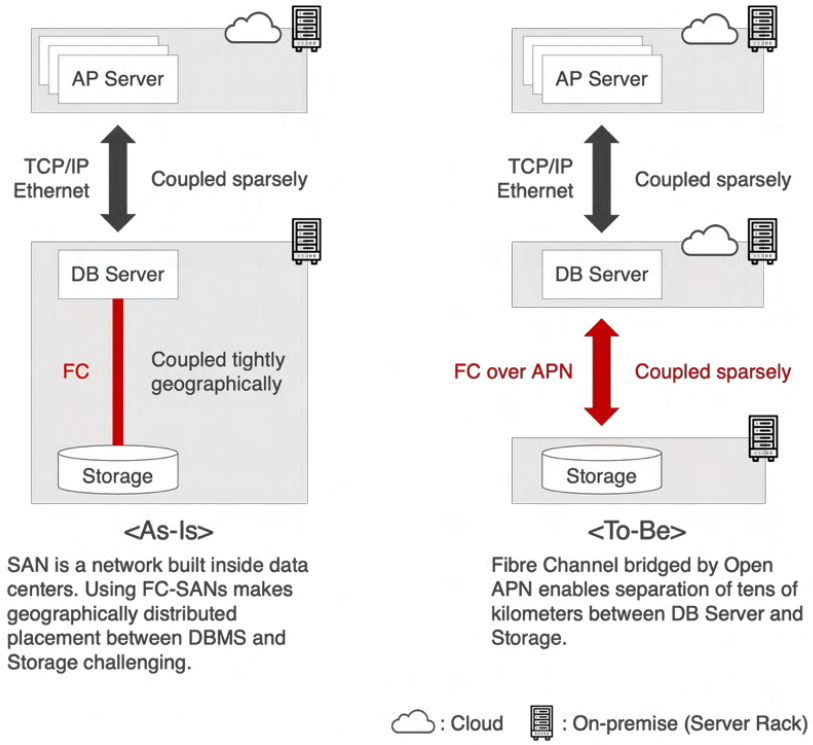

Figure A-1: Key point of FC-SAN over APN

A.3 Testing environment

As shown in Table A-1, Open APN and Fibre Channel utilize different light wavelength bands. Open APN leverages the C-band and L-band, which are optimal for long-distance, high-capacity transmission, whereas Fibre Channel uses the O-band.

As illustrated in Figure 5-2, it is necessary to use a transponder that supports Fibre Channel on the client-side interface and Open APN on the line-side interface. We searched for and adopted a transponder that meets these requirements. In addition, in this verification, the length between DC1.1 and DC2.1 was 70 km. To configure Open APNs of 90 km and 120 km, the distance shortfall was made up by means of a fiber-optic cable drum.

Table A-1: Optical communication wavelength band

Band name		Wavelength [nm]
T-band	Thousand-band	1000 ~ 1260
O-band	Original-band	1260 ~ 1360
E-band	Extended-band	1360 ~ 1460
S-band	Short-wavelength-band	1460 ~ 1530
C-band	Conventional-band	1530 ~ 1565
L-band	Long-wavelength-band	1565 ~ 1625
U-band	Ultralong-wavelength-band	1625 ~ 1675

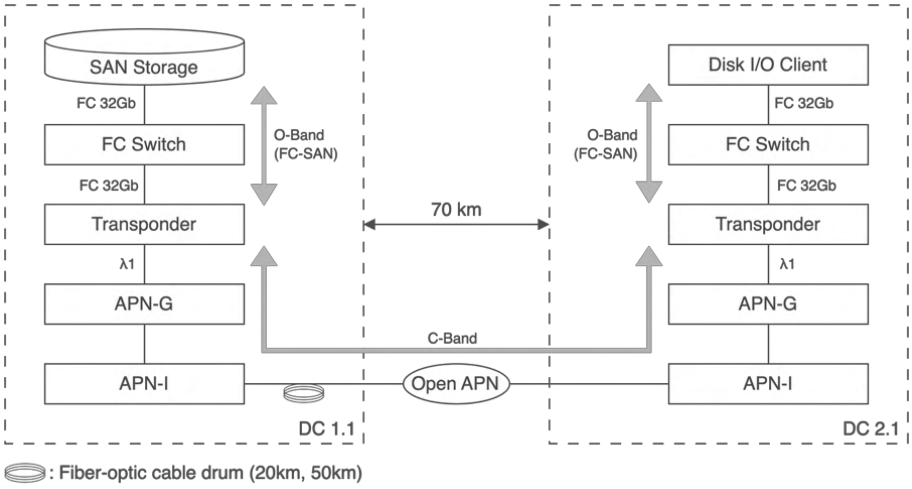

Figure A-2: Testing environment of FC-SAN over Open APN

A.4 Evaluation method and measurement

To accurately evaluate the performance impact of Open APN integration, benchmarks were performed using fio (flexible I/O tester v3.35), with each test lasting one minute. Results were recorded in terms of throughput—measured both by Input/Output Operations Per Second (IOPS) and bandwidth—and latency. The fio tests were configured in Synchronous I/O and Direct I/O modes to ensure disk access bypassed the file system cache for a more precise assessment of network effects.

The underlying SAN storage utilized the XFS file system, mounted with default options to maintain consistency across all measurements. Benchmark variables included the physical distance between data centers, specifically set at 70, 90, and 120 kilometers, to evaluate performance over varying spans of Open APN. The I/O workload targeted both sequential write and sequential read operations, with a fixed block size of 8 KB. Concurrent I/O was tested across a range of thread counts—1, 4, 8, 16, 32, and 64—to analyze scalability and stress different concurrency levels within the SAN environment.

A.5 Testing results

The measurement results of throughput and latency for sequential write are shown in Figure A-3. Even in the FC-SAN over Open APN configuration, it was confirmed that data can be written normally up to 120 km between DCs. It was confirmed that if the block size and the number of threads were the same, the throughput would decrease depending on the distance between the DCs. And it was confirmed that the latency of synchronous data writing also increased proportionately to the distance between DCs.

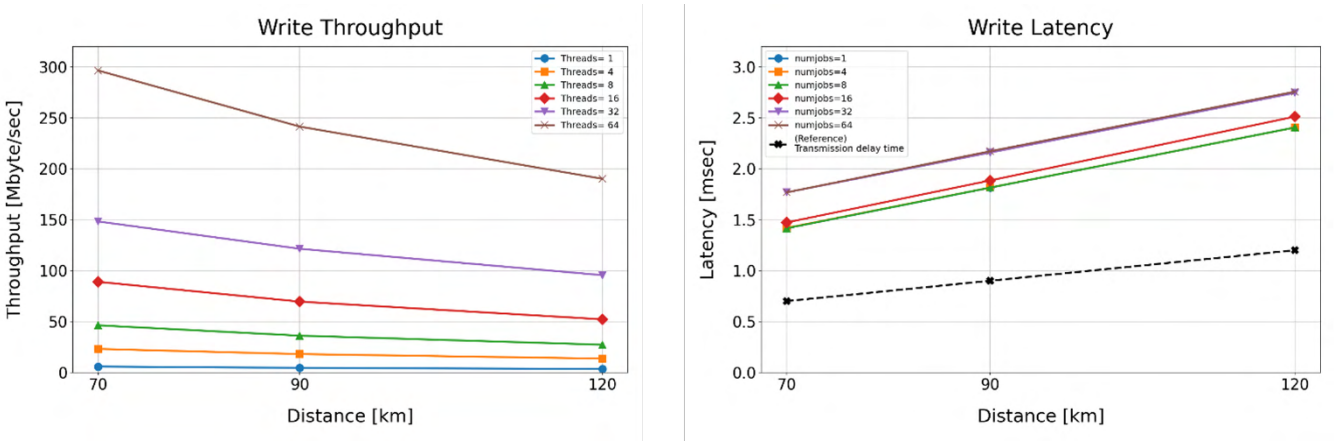

Figure A-3: Write throughput result(left) and write latency result(right)

The measurement results of throughput and latency for sequential write are shown in the FigureA-4. Even in the FC-SAN over Open APN configuration, it was confirmed that data can be read normally up to 120 km between DCs. Regarding the performance characteristics when the distance increases, they remain the same as those for Write operations.

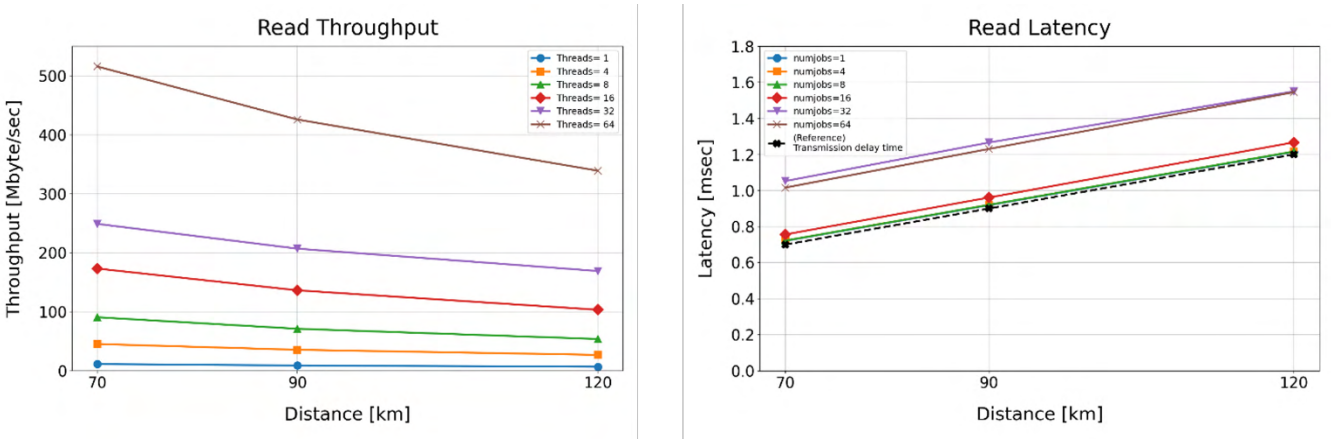

Figure A-4: Read throughput result (left) and read latency result(right)

A.6 Summary

When configuring an FC-SAN over Open APN with inter-DC distances ranging from 70 to 120 km, it was confirmed that read and write processes could be successfully performed. This achievement marks the first-ever interconnectivity within the system.

An example of methodology and equipment setup for building an FC-SAN across Open APN was established, followed by performing IO performance tests on remote storage systems. Current technologies and equipment require optoelectronic conversion because the frequency band used by Fibre Channel (O-Band) differs from those used by Open APN (C-Band, L-Band). In this test, an FC-compatible transponder was used for optoelectronic conversion, with the Client side using Fibre Channel and the Line side using Open APN. However, since the All-Photonics Network concept aims for end-to-end optical communication, the future development of wavelength conversion technologies between O-Band and C/L-Band will be necessary and important.

Through consortia like the IOWN Global Forum, feedback from use case teams should be communicated to the technology evaluation task force, fostering technological development and innovation in the field.

Appendix B. Supplementary specifications for PoC

This chapter describes the hardware specifications used in PoC 1 and PoC 2.

B.1 Hardware specifications of PoC 1

Table B-1: Hardware Specification of PoC 1

Component	Device Type		Specification
IA Server#1-3 (Virtualization Platform)	CPU		16cores
	Memory		128GB
	Disk		3680GB
IA Server#4 (Load Generator)	CPU		6cores
	Memory		64GB
	Disk		200GB

B.2 Hardware specifications of PoC 2

Table B-2: Hardware Specification of PoC 2

Component	Device Type		Specification
DB Server (Primary, Secondary)	CPU		2.8GHz 1P/32C
	Memory		128GB (32GB x 4)
	RAID Controller		12Gbps
	Disk		1.48TB (744.6GB x 2)
	Network	NIC	1GbE x 2port
		NIC	25GbE x 2port
		HBA	32Gb x 2port
	Others	IPMI	1port
Load Generator	CPU		2.9GHz 2P/16C
	Memory		512GB (32GB x 16)
	RAID Controller		12Gbps
	Disk		892GB (446.6GB x 2)
	Network	NIC	1GbE x 2port
		NIC	200GbE x 2port
	Others	IPMI	1port
SAN Storage	CPU		2.2GHz 2P/8C
	Memory		128GB
	Memory (NVMEM)		8GB
	Disk		43.56TB (3.63TB x 12)
			21.46TB (894GB x 24)
	Network	NIC	1GbE x 4port
		NIC	10/25GbE x 2
		HBA	32Gb x 4

History

Revision	Release Date	Summary of Changes
1.0	November 12, 2025	Initial Version