# Reference Implementation Model and Proof-of-Concept Reference of Services Infrastructure for Financial Industry Use Case

Classification: APPROVED REFERENCE DOCUMENT

Confidentiality: PUBLIC

Version 1.0

February 05, 2025

[FSI Use Case RIM PoC Ref]

# Legal

THIS DOCUMENT HAS BEEN DESIGNATED BY THE INNOVATIVE OPTICAL AND WIRELESS NETWORK GLOBAL FORUM, INC. (“IOWN GLOBAL FORUM”) AS AN APPROVED REFERENCE DOCUMENT AS SUCH TERM IS USED IN THE IOWN GLOBAL FORUM INTELLECTUAL PROPERTY RIGHTS POLICY (THIS “REFERENCE DOCUMENT”).

THIS REFERENCE DOCUMENT IS PROVIDED ”AS IS” WITH NO WARRANTIES WHATSOEVER, WHETHER EXPRESS, IMPLIED, STATUTORY, OR OTHERWISE, INCLUDING WITHOUT LIMITATION ANY WARRANTY OF MERCHANTABILITY, NONINFRINGEMENT OF THIRD PARTY RIGHTS, TITLE, VALIDITY OF RIGHTS IN, FITNESS FOR ANY PARTICULAR PURPOSE, OR ANY WARRANTY OTHERWISE ARISING OUT OF ANY PROPOSAL, REFERENCE DOCUMENT, SAMPLE, OR LAW. WITHOUT LIMITATION, IOWN GLOBAL FORUM DISCLAIMS ALL LIABILITY, INCLUDING WITHOUT LIMITATION LIABILITY FOR INFRINGEMENT OF ANY PROPRIETARY RIGHTS AND PRODUCTS LIABILITY, RELATING TO USE OF THE INFORMATION IN THIS REFERENCE DOCUMENT AND TO ANY USE OF THIS REFERENCE DOCUMENT IN CONNECTION WITH THE DEVELOPMENT OF ANY PRODUCT OR SERVICE, AND IOWN GLOBAL FORUM DISCLAIMS ALL LIABILITY FOR COST OF PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES, LOST PROFITS, LOSS OF USE, LOSS OF DATA OR ANY INCIDENTAL, CONSEQUENTIAL, DIRECT, INDIRECT, PUNITIVE, EXEMPLARY, OR SPECIAL DAMAGES, WHETHER UNDER CONTRACT, TORT, WARRANTY OR OTHERWISE, ARISING IN ANY WAY OUT OF USE OR RELIANCE UPON THIS REFERENCE DOCUMENT OR ANY INFORMATION HEREIN.

EXCEPT AS EXPRESSLY SET FORTH IN THE PARAGRAPH DIRECTLY BELOW, NO LICENSE IS GRANTED HEREIN, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS OF THE IOWN GLOBAL FORUM, ANY IOWN GLOBAL FORUM MEMBER OR ANY AFFILIATE OF ANY IOWN GLOBAL FORUM MEMBER. EXCEPT AS EXPRESSLY SET FORTH IN THE PARAGRAPH DIRECTLY BELOW, ALL RIGHTS IN THIS REFERENCE DOCUMENT ARE RESERVED.

A limited, non-exclusive, non-transferable, non-assignable, non-sublicensable license is hereby granted by IOWN Global Forum to you to copy, reproduce, and use this Reference Document for internal use only. You must retain this page and all proprietary rights notices in all copies you make of this Reference Document under this license grant.

THIS DOCUMENT IS AN APPROVED REFERENCE DOCUMENT AND IS SUBJECT TO THE REFERENCE DOCUMENT LICENSING COMMITMENTS OF THE MEMBERS OF THE IOWN GLOBAL FORUM PURSUANT TO THE IOWN GLOBAL FORUM INTELLECTUAL PROPERTY RIGHTS POLICY. A COPY OF THE IOWN GLOBAL FORUM INTELLECTUAL PROPERTY RIGHTS POLICY CAN BE OBTAINED BY COMPLETING THE FORM AT: www.iowngf.org/join-forum. USE OF THIS REFERENCE DOCUMENT IS SUBJECT TO THE LIMITED INTERNAL-USE ONLY LICENSE GRANTED ABOVE. IF YOU WOULD LIKE TO REQUEST A COPYRIGHT LICENSE THAT IS DIFFERENT FROM THE ONE GRANTED ABOVE (SUCH AS, BUT NOT LIMITED TO, A LICENSE TO TRANSLATE THIS REFERENCE DOCUMENT INTO ANOTHER LANGUAGE), PLEASE CONTACT US BY COMPLETING THE FORM AT: https://iowngf.org/contact-us/ Copyright © 2025 Innovative Optical Wireless Network Global Forum, Inc. All rights reserved. Except for the limited internal-use only license set forth above, copying or other forms of reproduction and/or distribution of this Reference Document are strictly prohibited.

The IOWN GLOBAL FORUM mark and IOWN GLOBAL FORUM & Design logo are trademarks of Innovative Optical and Wireless Network Global Forum, Inc. in the United States and other countries. Unauthorized use is strictly prohibited. IOWN is a registered and unregistered trademark of Nippon Telegraph and Telephone Corporation in the United States, Japan, and other countries. Other names and brands appearing in this document may be claimed as the property of others.

# Contents

# List of Figures

# List of Tables

# Executive Summary

The document “Reference Implementation Model and Proof-of-Concept Reference - Service Infrastructure for Financial Industry” defines two key components: the Reference Implementation Model and the Proof of Concept (PoC) Reference.

The Reference Implementation Model provides a practical implementation of IOWN technologies as a reference model to realize the service infrastructure for financial institutions.

The PoC Reference offers guidelines for conducting PoCs for the use case. It evaluates the Reference Implementation Model using defined technology evaluation criteria.

Two primary use cases are discussed in the document. It describes a common abstract system architecture to prepare for conducting PoCs and provides system sequence diagrams to precisely define the benchmark instrumentation for these use cases.

Those willing to conduct PoCs can select one of two use cases and combine multiple trials to create a concrete PoC.

# 1. Introduction

## 1.1. Purpose

To complete this RIM and PoC, we analyzed the challenges financial institutions face in the “Service Infrastructure for Financial Industry Use Case” document. Significant and ongoing shifts are occurring in digital banking services, including cross-industry collaboration/competition, leveraging data to provide a personalized customer experience, navigating and automating regulatory compliance, and maintaining legacy systems while delivering next-generation services.

The Service Infrastructure for Financial Industry project aims to deliver performance and reliability far beyond existing solutions, enabling financial entities to compete successfully in this crowded and highly regulated market. The Innovative Optical and Wireless Network Global Forum (IOWN GF), an association of the world’s most advanced technology leaders, has dedicated its collective efforts to developing new technologies that will enable the financial sector to address these challenges with agile and resilient services.

To fulfill its purpose, the project proposes that a financial service interconnect its own data centers and industrial, private, and public clouds to build and operate its own zone (like availability zones), allowing application deployment and workload migration. The service infrastructure leverages highly advanced IOWN technologies, such as Open All-Photonic Networks (Open APNs), which primarily rely on photonics (using laser light rather than electricity) as the basis for data transmission. This approach offers far higher energy efficiency, performance, resiliency, and more dynamic and flexible network management than existing technologies.

Eventually, the project's value is streamlined requirements and reduced operational expenditures, which provide a more cost-effective approach with a lower Total Cost of Ownership (TCO) and greater Return on Investment (ROI) compared to legacy data infrastructure solutions.

## 1.2. Project Objective

The objective of the Service Infrastructure for Financial Industry project aims to define a reference design for multi-data center computing infrastructure with the advanced capabilities that financial institutions require for their agile and resilient infrastructure.

The scope of this project is to:

1. Describe the Service Infrastructure for the Financial Industry Use Case and its key requirements.
2. Define the Technology Evaluation Criteria, which include reference cases and critical benchmarks.
3. Develop the Reference Implementation Model, which provides a practical implementation of IOWN technologies as a reference model to realize the use case.
4. Define the Proof of Concept (PoC) Reference, which provides guidelines for conducting PoCs for the use case to evaluate the Reference Implementation Model with the defined Technology Evaluation Criteria.
5. Develop and evaluate the PoC based on the Reference Implementation Model PoC Reference.

The primary objective of this document is to outline how IOWN technologies like Financial Services Infrastructure with IOWN will help financial services organizations develop game-changing services with agility while maintaining very high service resiliency with Financial Services Infrastructure with IOWN enhanced disaster recovery capabilities.

## 1.3. Document Scope

We have completed the first step of this activity, covering Steps 1 and 2. Please refer to the document “Use Case and Technology Evaluation Criteria—Service Infrastructure for Financial Industry” for further information. This document is the second step of this activity, which covers Steps 3 and 4 to engage early adopters in the financial industry.

The following sections of this document describe one of the Reference Implementation Models of Financial Services Infrastructure and define its system validation testing and performance benchmark as Proof-of-Concept references.

There are several approaches to building financial service infrastructure, such as container technologies; these will be addressed in future versions of this document.

# 2. Reference Implementation Model

In use cases related to DC-Inter Connect for financial institutions, we discuss two main points:

1. Application deployment and data migration between data centers to improve the robustness and agility of operations within a region.
2. Data backup and migration to improve the robustness of the system across regions.

The physical distance between data centers should be referenced to distinguish between Intra-Region and Inter-Region. Each is defined as follows in [IOWN FSI UC] of Appendix 4:

- **Intra-Region:** 50-100km
- **Inter-Region:** (Up to Geo-Location)

Note 1: The crucial point in the inter-region designation is each data center's independence. Each data center must have independent areas for shared services like Power, Water, and Gas, independent maintenance areas for data center support engineering, and be situated in different seismic regions. However, the distance ensuring this independence varies according to the Geo-Logical Condition. Therefore, this paper does not define a unified distance but sets distances for each Geo-Logical Condition in Appendix 3.

Note 2: The distance value regarding “Inter-Region” differs from the value described in the initial document [IOWN FSI UC]. Because the value of 500-1000 km may not always be appropriate as RIM, depending on the Geo-Location. Therefore, the initial document needs to be revised. The values regarding Inter-Region defined in [IOWN FSI UC] should be varied according to the definition in this 2nd document.

## 2.1. Basic Strategy of System Architecture and Design

Financial institutions face pressure from customers and markets to take advantage of IT innovations and offer digital “smart” services to help a digitally savvy “smart” society with financial technology.

We propose using IOWN GF technologies to solve financial institutions' issues. This would enable the synchronous operation of relational database management systems, replicating data from one data center to another. The architecture of backup systems will be significantly simplified, which can reduce operational costs in financial institutions' systems.

As a first step, we apply a design strategy leveraging IOWN networking technology—Open All-Photonic Networks (Open APN)—for data center and hybrid cloud interconnections. This is combined with commercially off-the-shelf, open-source, or proprietary virtual machine hosting platform technology or container-oriented application hosting platform technologies. With Open APN’s dynamic path allocation functionality and on-demand cost model, we reduce connectivity costs from traditional technologies.

*Table 1: Example of Technology Assumptions of each type of Platform Layers*

| Type of technology | Technology example |
|---|---|
| **Database Management System** | Relational DBMS, Distributed DBMS |
| **Operating System** | Linux |
| **Hosting Platform** | Virtual Machine hosting platform, Container hosting platform |
| **Hosting Hardware and interconnects** | IA Server, Smart NIC, PCI-Express over Ethernet |
| **Storage System** | Network-attached Storage system, Software-defined Storage system, Storage with NVMe over Open APN |
| **Local Area Network** | 10G/25G/100G Ethernet, TCP/IP, Photonics Network |
| **Wide Area Network** | Open APN, Deterministic Networking Service |

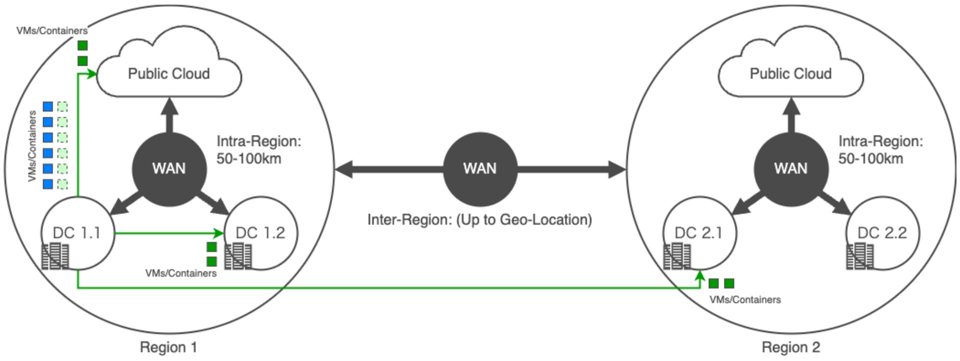


*Figure 1: Financial Data Center Inter-Connect*

## 2.2. Reference Implementation Model for Intra-Regional Application Deployment and Migration

Use Case 1, “Intra-regional Application Deployment and Migration,” realizes a flexible resource pool and application migration from one data center to another in the same regional area.

### 2.2.1. System Architecture

For the first use case, “Intra-regional Application Deployment and Migration,” the system operates in two types of locations: data centers and public clouds, assuming there are at least two data centers. Regarding system architecture, while we omit graphical representation, it should be noted that the physical distance between data centers or between a data center and a public cloud is approximately 50 to 100 kilometers.

Each data center has at least one server on which virtualization infrastructure is built, and at least one virtual machine is running. Databases and user applications are assumed to run within the virtual machines. Both virtual and physical machines have appropriate network interfaces and are connected to the emerging WAN through physical and virtual network devices. In addition, virtual servers exist in the public cloud, and it is assumed that they provide services equivalent to those running on servers in the data centers. The model assumes that data centers and public clouds are interconnected via WAN connectivity.

We assume that service platforms for financial institutions use virtual machine platforms as their common fundamental service, providing computer resources to applications. An important component is a Database Management System because financial services have the characteristics of “Record of Activity” systems.

Another critical component is storage. We can provide the storage service with dedicated storage equipment or software-defined storage.

These equipment or service nodes are placed in multiple data centers. For reference, we define the Primary Data Center (DC1.1) and the Secondary Data Center (DC1.2). We assume these data centers are owned or managed by financial institutions.

We also assume public cloud service. Computer resources, Storage services, and local networks, as defined, are provided by 3rd party service providers.

When performing VM migration, the way in which the virtualization infrastructure is built depends on the user's requirements. It is possible to configure a single virtualization infrastructure cluster across multiple data centers or to configure multiple virtualization infrastructures (in other words, a different virtualization infrastructure for each data center). When using a typical wide-area network or the internet for inter-DC connections, there are issues with throughput and latency, making it difficult to meet the Key Requirements defined in the former white paper. However, leveraging WAN is expected to solve these issues. Given that this project aims to discuss the ideal service infrastructure for innovative financial use cases using IOWN, this document assumes that a single virtualization infrastructure cluster will be configured across multiple data centers.

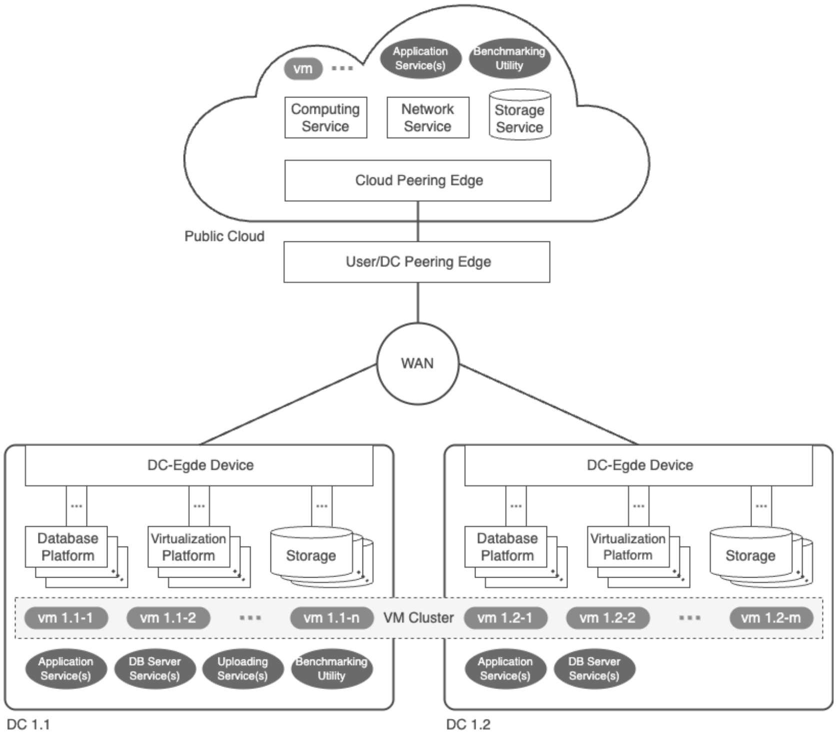


*Figure 2: Reference System Architecture for Use Case 1*

## 2.2.2. Functional Nodes

Table 2 describes the essential functional nodes in the reference implementation model. The functional nodes are logical components that run on bare-metal or virtualized server(s).

*Table 2: Node Profiles of Use Case 1*

| Node | Description | Attributes |
|---|---|---|
| **Database Platform** | This node accepts connections from applications run on the Virtualization Platform. DBMS software can be run on a virtual machine platform. Data storage can be on bare metal SSD or distributed software-defined storage service on a virtual machine platform. | • Place: DC<br>• # of Node: >1 per DC<br>• Interface: Ethernet<br>• # of Interface: 1> per Node |
| **Virtualization Platform** | This is a Virtual Machine host to provide computing resources to application services. The application services consist of multiple VM instances. The application services on each instance will communicate with each other. So, the platform provides a networking service for the application services.<br>The platform may host multiple guest virtual machines or container guests. VM hosts provide storage services to their guests, which can be local storage, distributed software-defined storage, or via network storage service which is provided by storage nodes. | • Place: DC<br>• # of Node: >1 per DC<br>• Interface: Ethernet, >25GbE<br>• # of Interface: 2> per Node |
| **Storage** | Storage nodes provide a block storage service to attach to VM instances by Virtualization Platform. The storage nodes can be realized with dedicated storage equipment, or software-defined storage (SDS) service with multiple server nodes.<br>The storage nodes also may provide an object storage service to host VM images.<br>When provided as SDS, the storage node provides redundancy among servers. To keep data redundancy, there are an odd number of storage service nodes and more than 3 nodes to protect virtual machines and files safety. | • Place: DC<br>• # of Node: 1> per DC<br>• Interface: iSCSI, NVMe over Fabric, or NVMe over TCP<br>• # of Interface: 1> per Node |
| **Computing Service** | This is a web service that allows users to utilize virtual computing resources. | • Place: Public Cloud<br>• # of Instance: >1 per Cloud<br>• Interface: Ethernet<br>• # of Interface: 1> per Node |
| **Network Service** | This is a web service that allows users to utilize virtual networking resources. | • Place: Public Cloud<br>• Interface: Ethernet<br>• # of Interface: 1> per Node |
| **Storage Service** | This is a web service that allows users to utilize virtual storage resources. | • Place: Public Cloud<br>• Interface: Ethernet<br>• # of Interface: 1> per Node<br>• Type of Storage: Block Storage, Object Storage |
| **Benchmarking Utility** | This tool gathers and monitors metrics and system performance to instrument the system. | |

### 2.2.3. System Sequence Diagrams

In this section, we describe the sequence diagram during VM migration execution. The sequence diagram focuses on the communication between elements and shows the series of processes from start to finish. The VM Migration use case defines the following scenarios. The sequence diagram and prerequisites for defining this use case are specified for each scenario.

**Scenario #1:** Intra-Region VM Migration from DC 1.1 to DC 1.2

**Scenario #2:** Intra-Region VM Migration from DC 1.1 to Public Cloud

#### 2.2.3.1. Scenario #1: Intra-Region VM Migration from DC 1.1 to DC 1.2

**Prerequisites:**

- The VM migration is performed as a Pre-Copy Live Migration.
- The source server and destination server are based on the same CPU architecture. This document does not cover discussions regarding CPU architecture differences.
- The VM migration process does not involve any human intervention.
- The destination server has sufficient hardware resources to run the migrated VM. This document does not cover discussions regarding resource allocation.

**Description:**

The downtime occurring during VM migration consists of the time required for each VM's Finalization ($T2$) and the aggregated downtime of these $T2$s ($SVC\text{-}T2$) (see note 1 below).

In Validation Testing, the KR is defined as having the $SVC\text{-}T2$ value be less than one second. This is because, regardless of how much the downtime of individual IT resources is reduced, the critical factor is the overall system's SLO, and the indicator that directly affects the SLO is the service's downtime. Therefore, to meet KR1.1, minimizing the service's downtime becomes important by either making the system components smaller or narrowing the scope of VM migrations conducted at one time.

Examples include segments in an Active/Standby redundant configuration, subsystems within the overall system, clusters of system components such as AP servers or Web servers, etc. Ultimately, ensuring the continuity of the service is crucial, and the evaluation indicator is having the $SVC\text{-}T2$ be less than one second.

> Note 1: Multiple functional clusters can set active-active or active-standby provisioning. Each application cluster has multiple virtual machines that typically serve as application servers, database management servers, or other roles. In this situation, downtime $SVC\text{-}T2$ is considered from when one of the virtual machines starts finalization to when all the VMs are resumed in the cluster.

**System Sequence Diagram:**

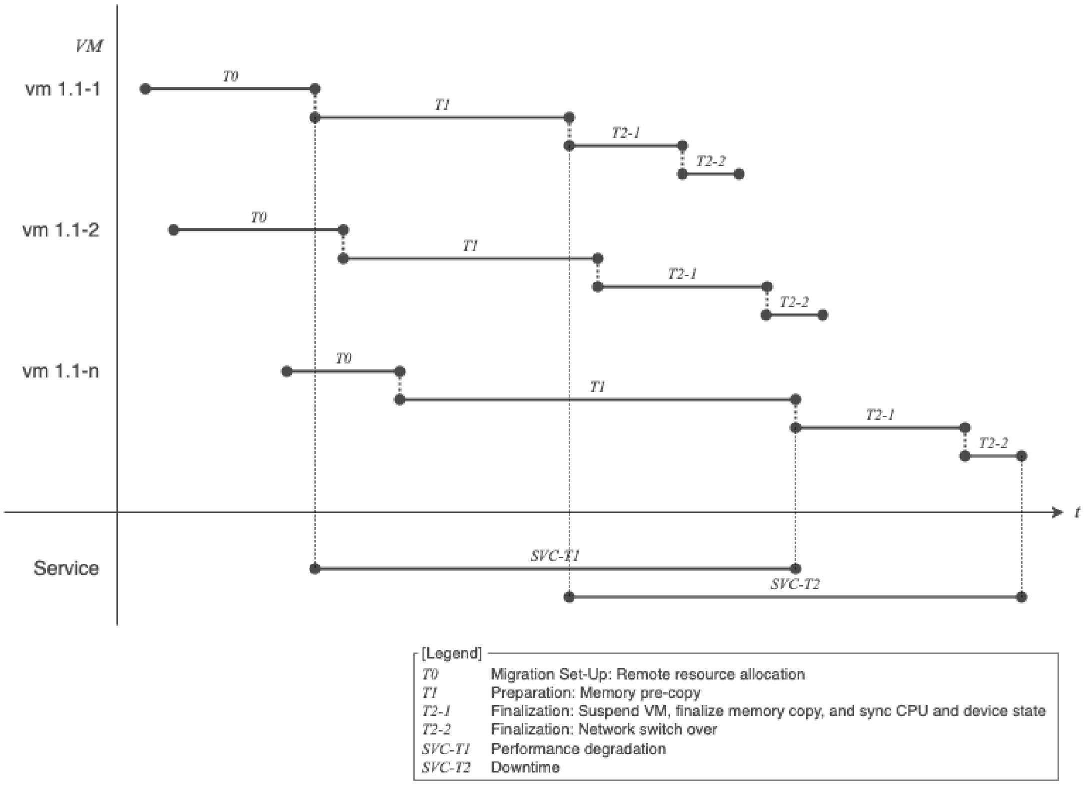


*Figure 3: Meta Sequence Diagram of VM Live Migration for Scenario #1 of Use Case 1*

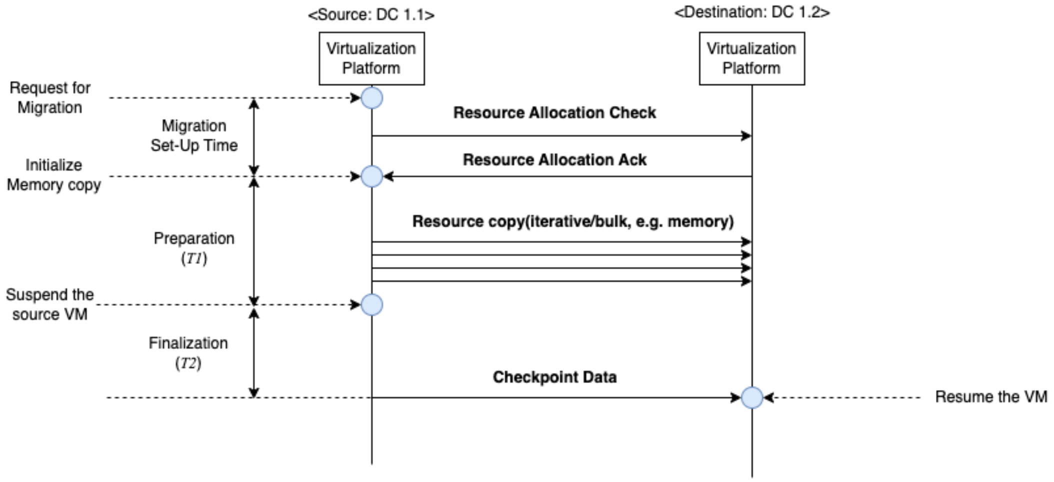


*Figure 4: Sequence Diagram of VM Live Migration for Scenario #1 of Use Case 1*

#### 2.2.3.2. Scenario #2: Intra-Region VM Migration from DC 1.1 to Public Cloud

**Prerequisites:**

- Separate virtualization infrastructure clusters must be constructed for the data center and the public cloud. The specific products (see note 2 below) should be used to ensure the virtualization infrastructure clusters are compatible with the data center and the public cloud for system portability.
- The target VM does not have any mount processes for using special hardware.
- The source server and destination virtual server are based on the same CPU architecture. This document does not cover discussions regarding CPU architecture differences.
- The destination virtual server has sufficient hardware resources to run the migrated VM. This document does not cover discussions regarding resource allocation.

**Description:**

Measure the time taken to copy the image of a virtual machine running on DC 1.1 to a public cloud and to start the VM on the public cloud. Calculate the total of these times to measure the overall duration for the virtual machine running on DC 1.1 to become operational on the public cloud.

**Scenario:**

- **Outputting the Image File:** Use the hypervisor functionality of DC 1.1 to output the target virtual machine to an image file. In this step, the entire virtual machine is saved to a file.
- **File Format Conversion (if necessary):** Convert the virtual machine image to a file format that can operate in the public cloud environment. Depending on the requirements of different virtualization technologies or cloud providers, this conversion process may be required.VM Image Transfer: The virtual machine image is copied and transferred to the public cloud's storage service via the APN (Advanced Packet Network). The transfer speed and bandwidth at this stage are also subject to time measurement.
- **VM Image Import:** In the public cloud, import the transferred VM image into storage. This process creates a new virtual machine (instance) from the VM image.
- **Applying Environmental Settings:** Other environmental settings (such as network and security settings) can be applied to the public cloud. This ensures that the virtual machine operates correctly in the public cloud environment.
- **VM Startup:** Start the newly created virtual machine on the public cloud. At this stage, confirm whether the virtual machine operates correctly.

**Times to be Measured:**

- **VM Image File Output Time:** The time it takes to output the virtual machine to an image file.
- **File Format Conversion Time (if necessary):** The time it takes to convert the image file to an appropriate format.
- **VM Image Transfer Time:** The time it takes to transfer the image file from DC 1.1 to the public cloud's storage service.
- **VM Image Import Time:** The time it takes to import the image file into the public cloud and create a new VM.
- **Environmental Settings Application Time:** Time taken to apply settings such as network and security configurations.
- **VM Startup Time:** Time taken to start the virtual machine.

**System Sequence Diagram:**

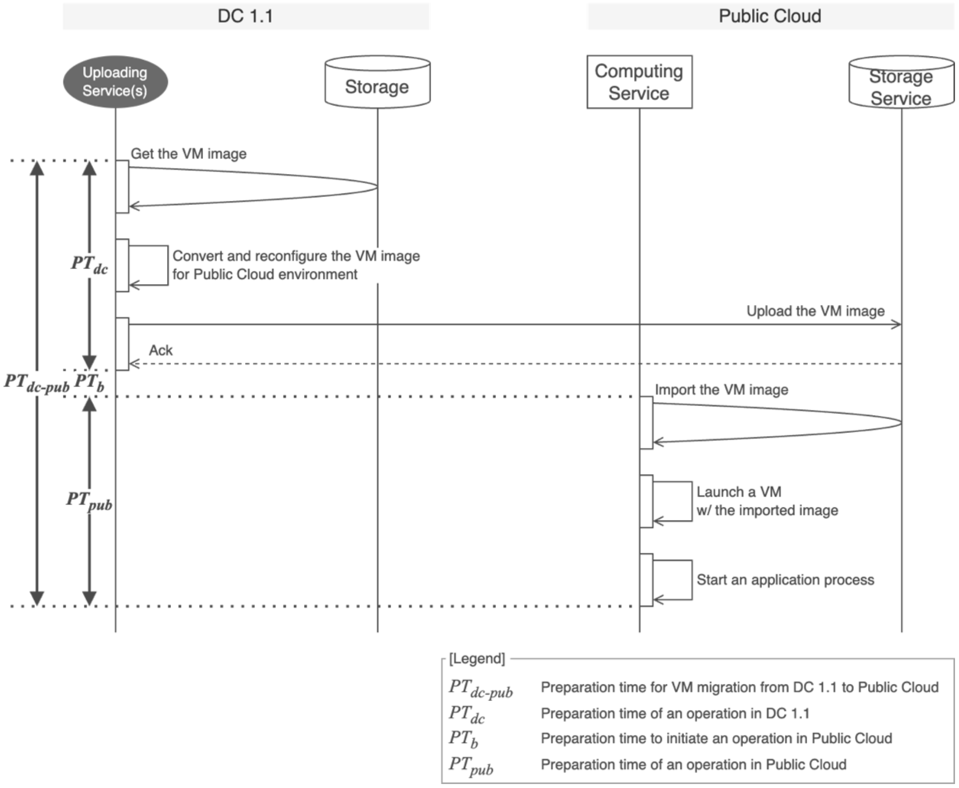


*Figure 5: Sequence Diagram of VM Migration for Scenario #2 of Use Case 1*

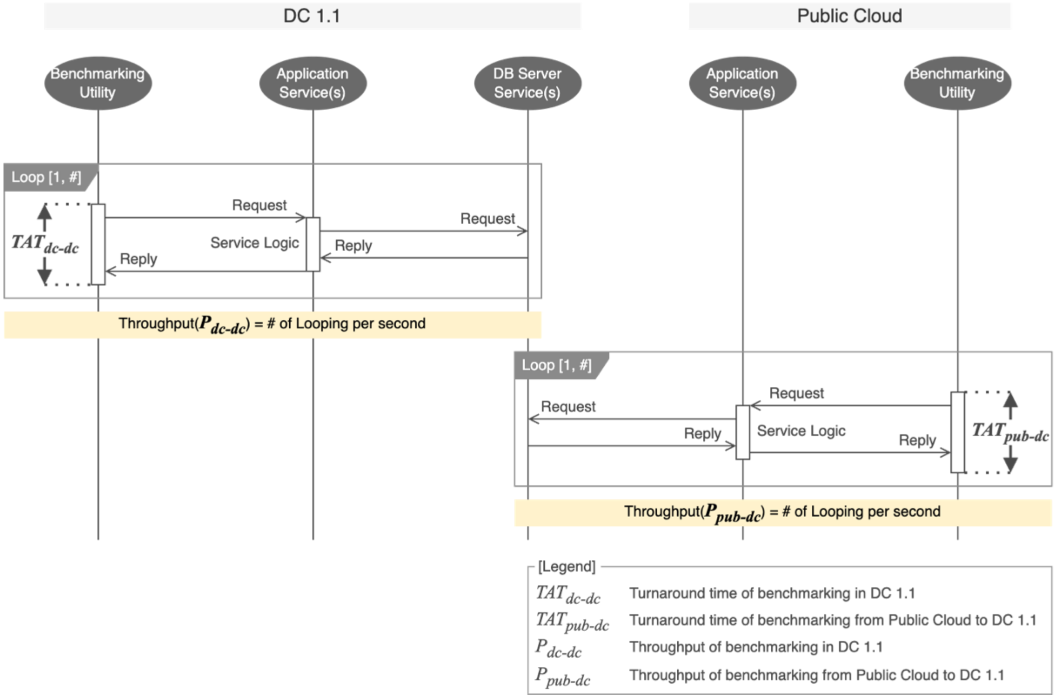


*Figure 6: Sequence Diagram of AP-DB Communication for Scenario #2 of Use Case 1*

Note 2: It is necessary to convert disk images when using a VM host platform with various software combinations, such as VMWare and Linux KVM/QEMU. Some software products support both private/on-premise environments and cloud environments with the same disk image. In this case, there is no need to prepare conversion software.

## 2.3. Reference Implementation Model for Inter-Regional Backups and Migration

The use case "Inter-regional Back-Ups and Migration to Improve Resiliency" connects two or more data centers in multiple cluster environments to improve resiliency against disaster or troubles by making back-up data and migrating applications to other data centers.

### 2.3.1. System Architecture

For the second use case, "Inter-regional Backups and Migration," the system operates in data centers, assuming there are at least two data centers.

From the perspective of network system architecture, the reference implementation model requires the WAN to operate over each minimum distance defined in Appendix 3. In contrast to the intra-region migration use case, which involves multi-site compute nodes within a single region stretch cluster designed for fault tolerance or high availability, the inter-region use case necessitates a multi-cluster design across the WAN to manage east-west traffic.

From the user systems' perspective, the reference implementation model is mainly assumed to be the same as the Intra-Region use case. However, a multi-cluster manager might be required to manage application deployments across the multi-cluster environment. Each cluster is designed as a fault-tolerant or high-availability system.

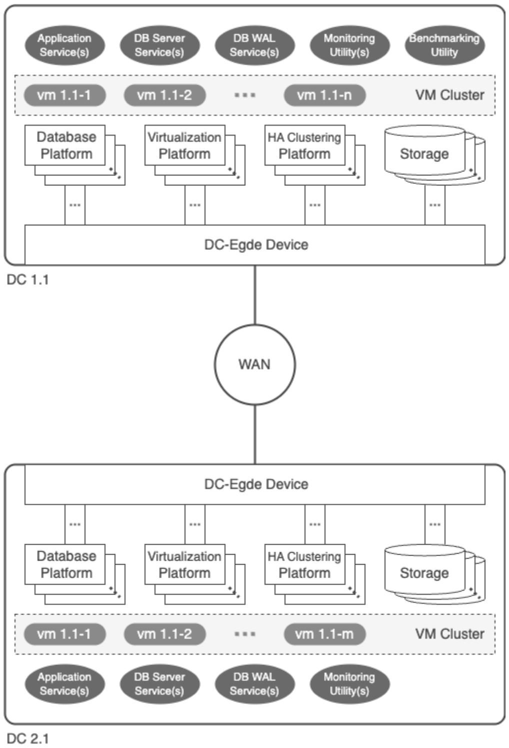


*Figure 7: Reference System Architecture for Use Case 2*

### 2.3.2. Functional Nodes

The functional nodes in this use case are almost identical to those in the intra-region use case. Therefore, describe only the functional nodes unique to the inter-region use case.

*Table 3: Node Profiles of Use Case 2*

| Node | Description | Attributes |
|---|---|---|
| **HA Clustering Platform** | A software solution that combines multiple nodes to provide a single virtual service with redundancy of nodes. It also monitors services to switch from a main service to a backup service when a problem happens.<br>This is necessary to achieve system redundancy with the backup center. The application process for HA clustering and the system node state monitoring process run on this platform.<br>The amount of traffic sent and received by this platform might not be high bandwidth, but it should be prioritized and controlled over the traffic of any other functional nodes. | • Place: DC<br>• # of Node: >1 per DC<br>• Interface: Ethernet<br>• # of Interface: 1> per Node |

### 2.3.3. System Sequence Diagrams

As stated in Section 2.3.3, the VM migration use case defines the following scenario. The prerequisites for determining the sequence diagram and its sequence are specified for each scenario.

**Scenario #1:** Inter-Region VM Migration

#### 2.3.3.1. Scenario #1: Inter-Region VM Migration

**Prerequisites:**

- The VM to be migrated does not involve any special mounting processes that require specialized hardware.
- The source server and the destination server share the same CPU architecture.
- The destination server has sufficient hardware resources to run the VM being migrated.

**Description:**

In the inter-region use case, VM migration is performed the same way as in the intra-region case. However, in the inter-region case, virtualization infrastructure is built in each data center. (The intra-region case assumes that a single virtualization infrastructure is built across multiple data centers) In the intra-region use case, it is assumed that a single virtualization infrastructure is built across multiple data centers; the sequence diagram related to VM migration itself is the same. For this reason, the sequence diagram for VM migration is omitted.

On the other hand, since the intra-region use case considers a disaster countermeasure scenario, data backup and synchronization between virtualization infrastructures are essential. Since different RTOs are required for each tier system, it is desirable to define economically rational data backup and synchronization processes for each tier system.

**Tier 1 System:**

Considering that it contains only about 1% of the total data volume, it is recommended to use the DB synchronous replication method. Generally, when using DB synchronous replication, data write performance to the database degrades compared to asynchronous replication, so this delay must be measured from an SLO perspective. Additionally, performance degradation is expected to be mitigated by IOWN Open APN.

**Tier 2 System:**

Considering that it contains about 10% of the total data volume, it is recommended to use either the DB synchronous replication method or the DB asynchronous replication method. Generally, when using DB asynchronous replication, the operator's workload increases during DR, and there may be a lag in data synchronization to the standby database. At this time, it is necessary to confirm that no user data has been lost (RPO=0).

**Tier 3 System:**

Considering that it contains about 90% of the total data volume, the recovery method from snapshots is utilized. The delay time from the start of data recovery to the service becoming listening again at the standby data center should be verified.

**System Sequence Diagram:**

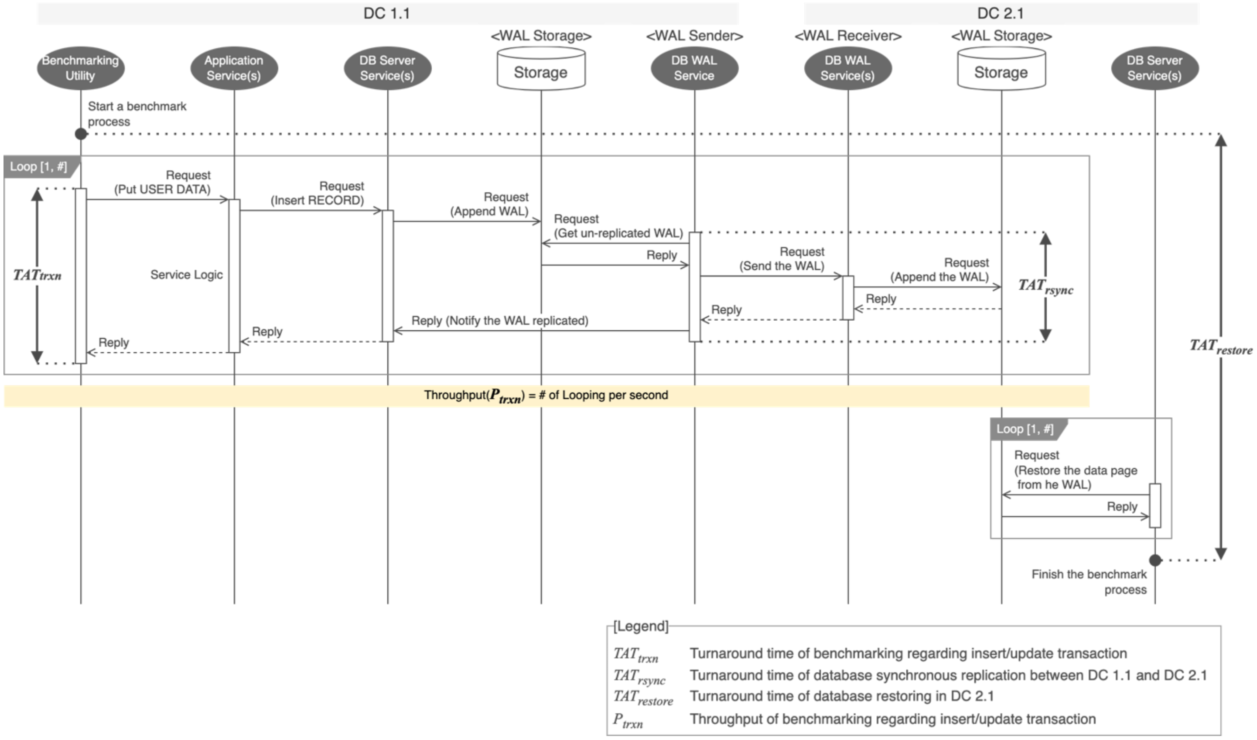


*Figure 8: Sequence Diagram of DB Synchronous Replication for Scenario #1 of Use Case 2*

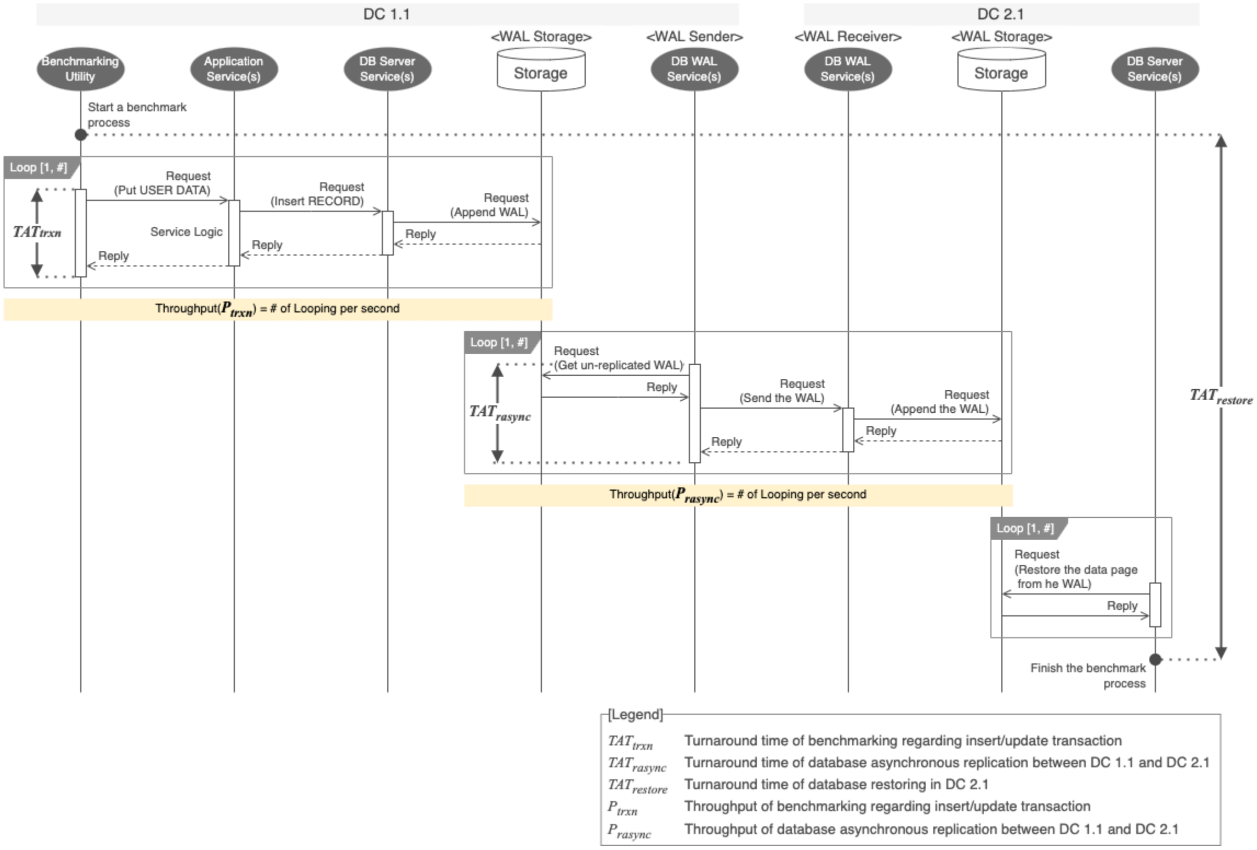


*Figure 9: Sequence Diagram of DB Asynchronous Replication for Scenario #1 of Use Case 2*

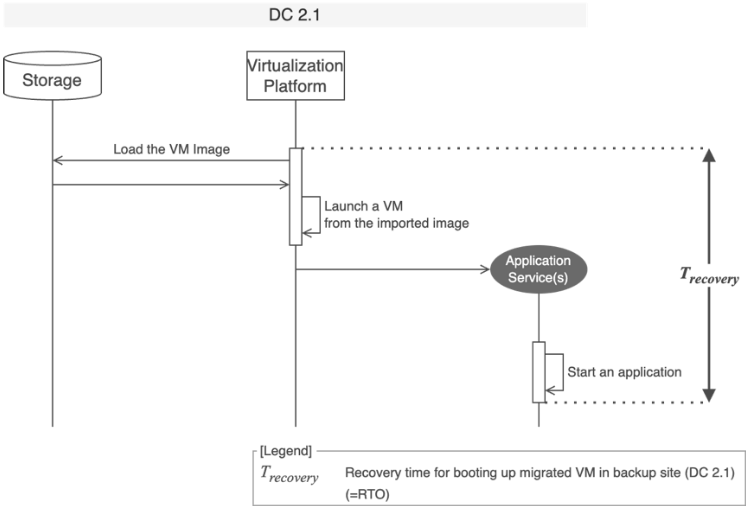
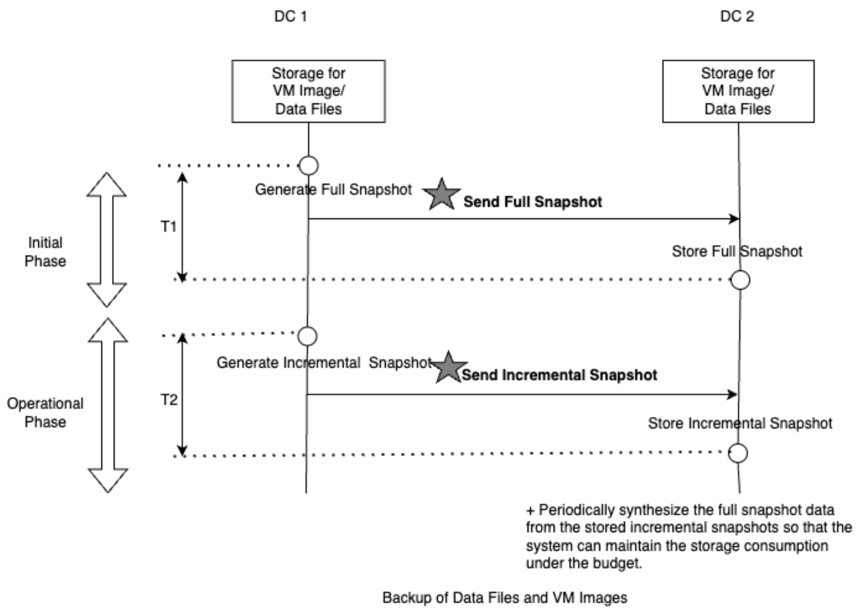


*Figure 10: Sequence Diagram of Recovering migrated VM for Scenario #1 of Use Case 2*



+ Periodically synthesize the full snapshot data from the stored incremental snapshots so that the system can maintain the storage consumption under the budget.

*Figure 11: Sequence Diagram of VM Snapshot for Scenario #1 of Use Case 2*

### 2.3.4. Handling DC 2.1 Failure When Synchronous Replication is Implemented

When synchronous replication is implemented, responses to clients will not be returned until the replication to DC 2.1 is complete. Therefore, it is necessary to consider measures to handle the failure of DC 2.1. The following approaches can be considered:

- **Prioritize Data Synchronization:** Design the system to accept the possibility that responses might not be returned to clients if data synchronization takes precedence.
- **Prioritize Continuation of DC 1.1 Processing:** Prepare a mechanism to temporarily switch to asynchronous replication to DC2 in the event of a failure.
- **Automation:** Since manual intervention by operators could take too long (resulting in prolonged waiting times for update transactions), it is necessary to introduce some form of automated response.

In detail:

- **System Design for Data Synchronization Priority:** If data synchronization is the top priority, design the system to tolerate the scenario where client responses might not be immediate. This approach ensures that data integrity is maintained above all else, even if it means delayed client interactions.
- **Temporary Switch to Asynchronous Replication:** If the priority is to ensure that DC 1.1 continues processing without interruption, implement a mechanism to automatically switch from synchronous to asynchronous replication in the event of DC 2.1 failure. This will allow DC1 to continue its operations and client interactions without dependency on DC 2.1.
- Automation for Swift Response: Introduce an automated system to handle the transition from synchronous to asynchronous replication in real-time. Manual intervention could lead to excessive delays, so an automated system can immediately respond to DC 2.1 failures, minimizing the wait time for updated transactions.

# 3. PoC Reference

In this section, we define PoC References for each of two use cases, intra- and inter-regional, using the technology evaluation criteria described in section 3 of [IOWN FSI UC] and the Reference Implementation Models defined in the previous section.

The reference cases in the technology evaluation criteria in 3.1.1 and 3.2.1 of [IOWN FSI UC] define the specific conditions that meet the key requirements for accurately evaluating the reference implementation model for the intra- and inter-regional use cases, respectively.

The PoC References in this section describe the methods for validating the key requirements and evaluating metrics for benchmarking. The key requirements are the objective criteria that need to be met in PoCs to realize financial use cases. Better metrics for benchmarking indicate that the PoC is performed with a more desirable implementation for users. It is important to note that the contents described here provide useful information on evaluating PoCs rather than established rules or regulations.

## 3.1. PoC Reference for Intra-Region Use Case

The Intra-Region use cases include scenarios for VM migration between data centers and VM migration from data centers to public clouds. The key requirement regarding the intra-region use case is below:

- **KR1.1:**
    - Downtime ($SVC\text{-}T2$) is less than 1[sec].

### 3.1.1. Scenario #1: Intra-Region VM Migration from DC 1.1 to DC 1.2

**Validation Testing:**

- **Prerequisites for Evaluation:**
    - The [Virtualization Platform], which is a component of the reference financial system, must have the ability to output metrics related to Preparation ($T1$) and Finalization ($T2$) that occur during VM Migration.
    - The time taken for network switchover, such as address re-resolution between MAC/IP and MAC table updates within network devices after VM migration, can be treated as 0 seconds. Although it is part of the VM migration process, it is often less valuable than the time required for other computational processes like memory copying.
- **Metrics:**
    - Start time, End time, and Processing time:
        - Finalization ($T2$) for all migrated VMs
        - Downtime ($SVC\text{-}T2$)
- **Evaluation Methods:**
    - Perform migration of multiple VMs either simultaneously or sequentially.
- **Points to Verify:**
    - It is quite possible for the execution times of Preparation ($T1$) and Finalization ($T2$) between VMs to overlap. Therefore, when calculating Downtime ($SRV\text{-}T2$), care must be taken not to double-count these times.
    - The load on the group of VMs targeted for migration is not stipulated as a condition. However, measurements should be taken while the system runs close to levels seen during actual use or under

continuous server/client load. Therefore, it is necessary to quantitatively understand the load on the VM group during the evaluation.

**Benchmark Testing:**

- **Prerequisites for Benchmark:**
  - (Same as Validation Testing)
- **Metrics:**
  - Start time, End time, and Processing time:
    - Preparation time (*T1*) for all migrated VMs
    - Performance Degradation time (*SVC-T1*)
- **Benchmark Methods:**
  - (Same as Validation Testing)
- **Points to Benchmark:**
  - Continuity of processes at the application layer (whether processes continue after VM migration is executed).
  - Consistency of data held by the database (whether there are no inconsistencies in data registration before and after VM migration execution).
  - Application performance degradation (e.g., Response time / Turnaround time, Number of errors, and Number of transactions)

### 3.1.2. Scenario #2: Intra-Region VM Migration from DC 1.1 to Public Cloud

**Validation Testing:**

- **Remarks:** The [IOWN FSI UC] does not define the key requirements for this scenario. However, downtime is an important metric that significantly impacts service levels and must be measured. Therefore, arbitrary reference values will be used in future PoCs. When the user performs a PoC, it should be confirmed that the value is better than this reference value, following the implementation procedures and perspectives defined in this Validation Testing.

**Benchmark Testing:**

- **Prerequisites for Benchmark:**
  - **DC Environment:**
    - The migration targets are multiple virtual machines (VMs) on a hypervisor.
    - The OS and software of the VMs must be compatible to run on the public cloud.
    - The VMs to be migrated should be convertible to a data format that cloud services can import.
    - The output of VM images in the DC should follow the hypervisor's capabilities.
    - Devices mounted on the VMs are not to be migrated and should not affect the operation post-migration.
  - **Public Cloud Environment:**
    - The migration to the public cloud should be done sequentially.
    - The public cloud account has already been created, and the necessary permissions have been granted.
    - Authentication and API keys for accessing the public cloud have already been obtained.

    - The public cloud should have enough free resources to start the VMs.
  - **Between DC and Public Cloud:**
    - The connectivity between the DC and the public cloud should be feasible.
    - The DC and public cloud should have synchronized time.
- **Metrics**
  - Preparation Time ($T1$): From starting the VM image extraction in the DC until the VM starts running in the cloud.
  - Turnaround Time ($TATdc\text{-}dc$, $TATpub\text{-}dc$): These are defined in Figure 6.
  - Throughput ($Pdc\text{-}dc$, $Ppub\text{-}dc$): These are defined in Figure 6.
- **Benchmark Methods**
  - Logs from the hypervisor with the function to extract VM images.
  - Execution logs from the cloud service for starting the VMs.
- **Benchmark Procedure**
  1. Use the hypervisor's feature in DC1.2 to output the target virtual machines as image files.
  2. If necessary, convert the VM images into a file format that can operate in the cloud environment.
  3. Copy and transfer the VM images to the public cloud's storage service via APN.
  4. In the public cloud, import the VM images transferred to the storage and create the VMs (instances).
  5. Configure other environment settings (such as network and security groups) according to the public cloud environment.
  6. Start the created VMs and launch the application processes.

## 3.2. PoC Reference for Inter-Region Use Case

The Inter-Region use cases include scenarios of VM migration between data centers. The key requirements regarding the inter-region use case are below:

- **KR2.1:**
  - RPO = 0 (User data on system use for financial services cannot be lost.)
- **KR2.2:**
  - RTO = [Depends on kinds of system]
    - **Tier 1 System:**
      - RTO < 30[mins]
      - Additional time is acceptable if a human operator needs to restore unprocessed transaction data or check the consistency of data.
    - **Tier 2 System:**
      - RTO < 2[hours]
      - The backup period is not always real-time. Hourly or daily backup is ordinally timing for full backup.
    - **Tier 3 System:**
      - RTO < 24[hours]

    - The backup period is not always real-time. Daily or special maintenance time (e.g., holidays) is also ordinarily for full backup.

### 3.2.1. Scenario #1: Inter-Region VM Migration from DC 1.1 to DC 2.1

**Validation Testing:**

- **Prerequisites for Evaluation:**
    - The [Application VM/Container Platform], which is a component of the reference financial system, must have the ability to output metrics related to Preparation ($T1$) and Finalization ($T2$) that occur during VM Migration.
    - It is quite possible for the execution times of Preparation ($T1$) and Finalization ($T2$) between VMs to overlap. Therefore, when calculating Downtime ($SRV\text{-}T2$), care must be taken not to double-count these times.
    - The time taken for network switchover, such as address re-resolution between MAC/IP and MAC table updates within network devices that occur after VM migration, can be treated as 0 seconds. Although it is part of the VM migration process, it is often less valuable than the time required for other computational processes, like memory copying.
    - VM migrations may be live migrations.
    - The load on the group of VMs targeted for migration is not stipulated as a condition. However, measurements should be taken while the system runs at levels close to those seen during actual use or under continuous server/client load. Therefore, it is necessary to quantitatively understand the load on the VM group during the evaluation.
- **Metrics:**
    - Replication Lag (=RPO) = $TATrasync$
        - $TATrasync$: This is defined in Figure 9
    - Number of Unreplicated Transactions (Grey Data) = $Ptrxn * TATrasync$
        - $Ptrxn$, $TATrasync$: These are defined in Figure 9
- **Evaluation Methods:**
    - Conduct and validate both synchronous and asynchronous replication scenarios to ensure the robustness and reliability of data replication strategies.
- **Points to Verify:**
    - **For Asynchronous Replication:**
        - **Stable Replication Lag:** Ensuring stable and acceptable levels of replication lag.
        - **RTO Verification:** Confirming that RTO is achieved within the expected time, even in the presence of network-induced delays.
        - **Operational Procedures for RPO=0:** It establishes and verifies operational procedures or application implementations that ensure RPO=0. This includes measures to prevent client data loss during asynchronous replication and to maintain data integrity post-DR.
    - **For Synchronous Replication:**
        - **Data Integrity:** Ensuring that clients do not submit the same data operation multiple times, thereby preventing data replication anomalies.
        - **RPO=0:** Confirming that RPO=0 is inherently maintained by synchronous replication mechanisms.

**Benchmark Testing:**

- **Prerequisites for Evaluation:**
  - (Same as Validation Testing)
- **Metrics:**
  - Downtime of VM migration from DC 1.1 to DC 2.1:
    - $Trecovery$: This is defined in Figure 10
  - DB Performance, in Transaction / Sec (for Tier 1):
    - $TATtrxn$: This is defined in Figure 8 and Figure 9
    - $TATrsync$: This is defined in Figure 8
    - $TATrasync$: This is defined in Figure 9
    - $Ptrxn$: This is defined in Figure 8
    - $Prasync$: This is defined in Figure 9
  - Backup speed (for all Tiers):
    - $T1 + T2$: These are defined in Figure 11

  Note 1: It should ideally measure both the metrics of $TATtrxn$ and $Ptrxn$. However, if a DBMS that the user chooses for the user's PoC cannot support the output of the metrics regarding $TATtrxn$, the user may calculate $TATtrxn$ as "$TATtrxn = 1 / Ptrxn$". With the same approach, $TATrasync$ may be calculated as "$TATrasync = 1 / Prasync$".

- **Benchmark Methods:**
  - (Same as Validation Testing)

# 4. Conclusion

This document explains the basic system architecture and design strategy that leverages IOWN GF Technology to efficiently realize the Financial Service Infrastructure Use Case.

As explained in Section 1.3, this document covered 3 and 4 described in Section 1.2 Project Objective of engaging early adopters in financial institutions. We hope Proof-of-Concept demonstrations and evaluations will be conducted to prove the validity and effectiveness of IOWN GF Technology following this document.

# Appendix 1. Abbreviation

*Table 4: Abbreviation*

| Abbreviation | Meaning |
|---|---|
| **AP** | Application |
| **APN** | All-Photonics Network |
| **DB** | Database |
| **DBMS** | Database Management System |
| **DC** | Datacenter |
| **DCI** | Data-Centric Infrastructure |
| **IaaS** | Infrastructure as a Service |
| **IOWN** | Innovative Optical and Wireless Network |
| **IOWN GF** | IOWN Global Forum |
| **KR** | Key Requirements |
| **NIC** | Network Interface Card |
| **NVMe** | Non-Volatile Memory Express |
| **Open APN** | Open All-Photonic Network |
| **PoC** | Proof-of-Concept |
| **ROI** | Return on Investment |
| **RPO** | Recovery Point Objective |
| **RTO** | Recovery Time Objective |
| **SDS** | Software-Defined Storage |
| **SLO** | Service Level Objective |
| **TCO** | Total Cost of Ownership |
| **TPS** | Transaction per Second |
| **VM** | Virtual Machine |
| **VMM** | Virtual Machine Monitor |
| **WAL** | Write Ahead Log |

# Appendix 2. Glossary

*Table 5: Definition*

| Term | Definition |
|---|---|
| **Container** | A discrete environment is set up within an operating system in which one or more applications may be run. |
| **Hypervisor** | Also known as a VMM. A type of computer software or firmware that creates and runs VMs. |
| **Inter-regional** | Between or involving two or more different regions |
| **Intra-regional** | Within a region |
| **Live migration** | Also known simply as migration, refers to the process of moving a running VM or application across different physical machines without disconnecting the client or application. During this process, the memory, storage, and network connectivity of the VM are transferred from the original host machine to the destination machine. |
| **NVMe over Fabrics** | NVMe block storage protocol through storage networking fabrics, such as Fiber Channel, RoCE, and InfiniBand. |
| **Region** | A geographical area within a 50km radius, i.e. 100km in maximum distance (see appendix A of [IOWN FSI UC]). This comes from the geographical measures in Japan, but different values can be applied to other countries. |
| **RoCE** | RDMA over Converged Ethernet |

# Appendix 3. Distance of Inter-Region for each Geo-Locations

This table describes the minimum and maximum distances between inter-regions defined for each Geo-Location where the IOWN GF Member Meeting has been held in the past. The distance values are measured using public mapping services [OSM]. The minimum distance represents the distance between the cities where major data centers are located [AZURE DC] [GOOGLE DC] [AWS DC] [ORACLE DC], and the maximum distance is the longest distance between major cities in each Geo-Location.

ⓘ Note: It is essential to meet the minimum distance, not the maximum distance, to obtain endorsement from the IOWN GF when conducting a reference PoC.

*Table 6: Distance of Inter-Region for each Geo-Locations*

| Area | Geo-Location | Min. distance | Max. distance |
|---|---|---|---|
| **Asia** | Japan | 250 | 2,500 |
| | Taiwan | 125 | 300 |
| **Europe** | Germany | 125 | 800 |
| **North America** | Canada | 250 | 4,000 |
| | United States | 250 | 4,000 |

# Appendix 4. Reference

[AWS DC]: Global Infrastructure regions and AZ

[AZURE DC]: Azure global infrastructure experience

[GOOGLE DC]: Discover our data center locations

[IOWN FSI UC]: IOWN Global Forum, "Services Infrastructure for Financial Industry Use Case Version 1.0," 2024.

[ORACLE DC]: Public Cloud Regions and Data Centers | Oracle

[OSM]: OpenStreetMap

# History

| Revision | Release Date | Summary of Changes |
|---|---|---|
| 1.0 | February 05, 2025 | Initial version |