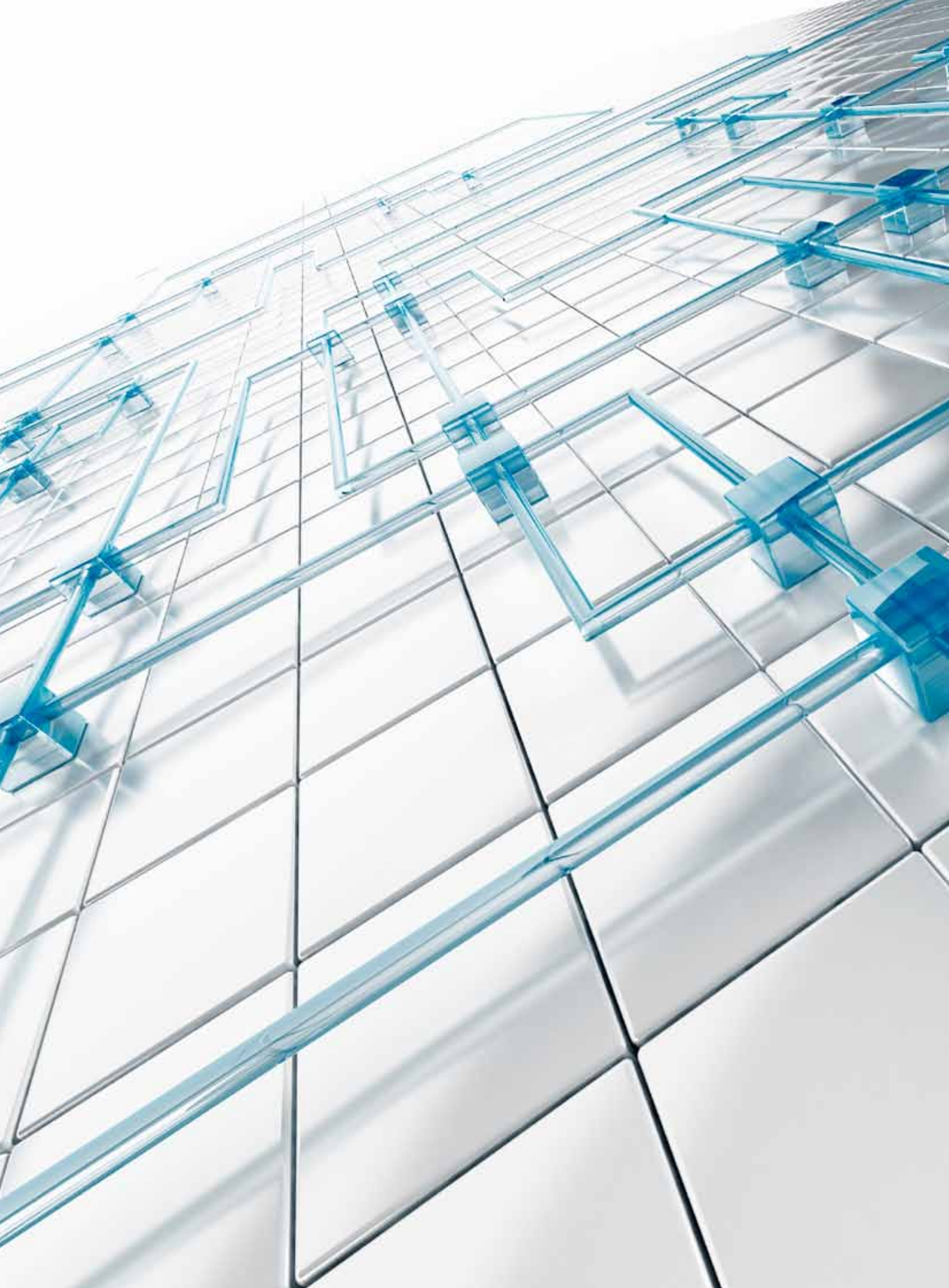


**Second (Nth)**

**Data Center High Availability Solution**



**HUAWEI**



# TABLE OF CONTENTS

|   |    |
|---|----|
| <b>DATA</b> -----   | 1  |
| <b>Executive Summary</b> -----                                  | 3  |
| <b>Introduction</b> -----                                       | 4  |
| <b>Challenges</b> -----   | 5  |
| <b>Solution</b> -----   | 6  |
| <b>Typical Scenarios</b> -----                                  | 8  |
| Active-Standby Data Centers-----                                | 8  |
| Summary-----  | 9  |
| Single Database Instance Deployed in Multiple Data Centers----- | 10 |
| Multiple Databases Deployed in Multiple Data Centers-----       | 12 |
| Summary-----  | 16 |
| <b>Solution Summary</b> -----                                   | 17 |



## Executive Summary

---

The traditional method has been deployment of service continuity system with the same architecture as the production system, using the system once a year for system testing or bring it online only during a data crisis, which means that utilization of the second data center remains very low. More importantly when a disaster occurs and the service continuity and disaster recovery systems are needed, there are a series of complicated and time-consuming steps, such as network switchover, mounting the storage, initiating the operating system, and starting applications, that lengthen system downtime and produce a number of dangers which could threaten the switchover with failure. These factors are not in line with the new and rigorous requirements that enterprises have for high IT application system usability, service continuity, and disaster recovery.

Since IT application systems become increasingly important to enterprise development, their continuity must be safeguarded by reliable high-availability solutions. That means upon a disaster, IT application systems must be taken over by another data center to maintain their business continuity.

For these reasons, Huawei and Oracle jointly released the Huawei Heterogeneous Active-Active Data Center Solution for Business Continuity and Disaster Recovery, aiming to significantly improve the utilization of storage system resources.



# Introduction

---

## Preface

There are many ways to achieve high availability and the methods outlined in this white paper are just some. Data consistency has always been a major focus for remote disaster recovery (DR) centers, and so this white paper will begin with by exploring this central challenge. This document will then detail high-level solutions and their theoretical bases.

## Overview

A growing number of enterprises tend to deploy their application systems in two or more data centers for high data availability and resource utilization rate. These data centers can be thousands of miles apart. The second (or Nth) data center can be a backup that synchronizes data with the primary data center in real time, as well as an independently functioning unit that processes local production transactions. Whether a second data center is qualified depends on its recovery time objective (RTO) and recovery point objective (RPO), and also its DR organizational capability such as finances, competitiveness, and construction time. HA is not just a technical issue, but requires a well-organized management. For most enterprises, either a short unexpected or planned downtime would result in huge losses, greater than the cost of building one or more data centers.

## When multiple data centers are deployed, the solution for business continuity and reliability must also possess the following:

**Resilient fault tolerance:** A site failure may be caused by natural disasters (fires, tornadoes, earthquakes), human misoperations (unintentional removal of cables, incorrect configurations), device abnormality (networks, databases, storage devices), or operating of only one data center at a time. To minimize the risk of failure, the solution must reserve a sufficient distance between data centers and interconnect them based on the methods described in this document, enabling enterprises to isolate points of failure for business continuity.

**High performance:** The closer a data center is to users, the higher performance the data center can provide. This philosophy has been widely applied in content delivery networks (CDNs). For example, average access latency between Japan and the UK over the Internet is several hundreds of milliseconds. In contrast, the latency of accessing local data centers in UK can be shortened to dozens of milliseconds. Performance may be notably improved when multiple data centers in different geographical locations are available for users to choose.

**Easy code releasing and maintenance:** When an enterprise deploys two or more self-owned data centers, it will be simpler to release code and conduct maintenance, shortening planned system downtime. If an enterprise upgrades data centers one by one and always keeps at least one data center running all the time, planned system downtime can be greatly shortened or even eliminated.

# Challenges

The deployment of multiple data centers faces three challenges:

**1.Data consistency:** Because of the latency in a distributed environment, the updates of the same record are not always synchronous in two or more databases located far away from each other. This asynchronous updates can cause data to be mistakenly overwritten, resulting in data unreliability.

For example, if two or more users use different terminals, such as laptops and smartphones, but the same account to log in to two data centers at the same time. In such a circumstance, these users share the same profile and perform the same operations on database records. No problems occur when the data centers are deployed in active-standby mode because users can write data in only one database. However, data may become inconsistent when the data centers are deployed in multi-active mode and the logged users with the same account are in different regions, countries, or continents. Even though it rarely happens, data inconsistency is unacceptable to most enterprises.

**2.Data synchronization performance:** Generally, there is a certain distance between the primary system and a DR system. Synchronizing data between them requires sufficient local data performance and network performance. If the primary system and DR system reside in the same cluster, a high-performance and low-latency system environment is required to arbitrate data consistency and take over sessions.

**3.Utilization of existing devices:** Since multiple data centers are always constructed at different times, the architectures and models of their servers, storage devices, networks, operating systems, and platform software may come from varying vendors. To ensure smooth data replication between data centers, a solution must enable heterogeneous systems to interoperate.

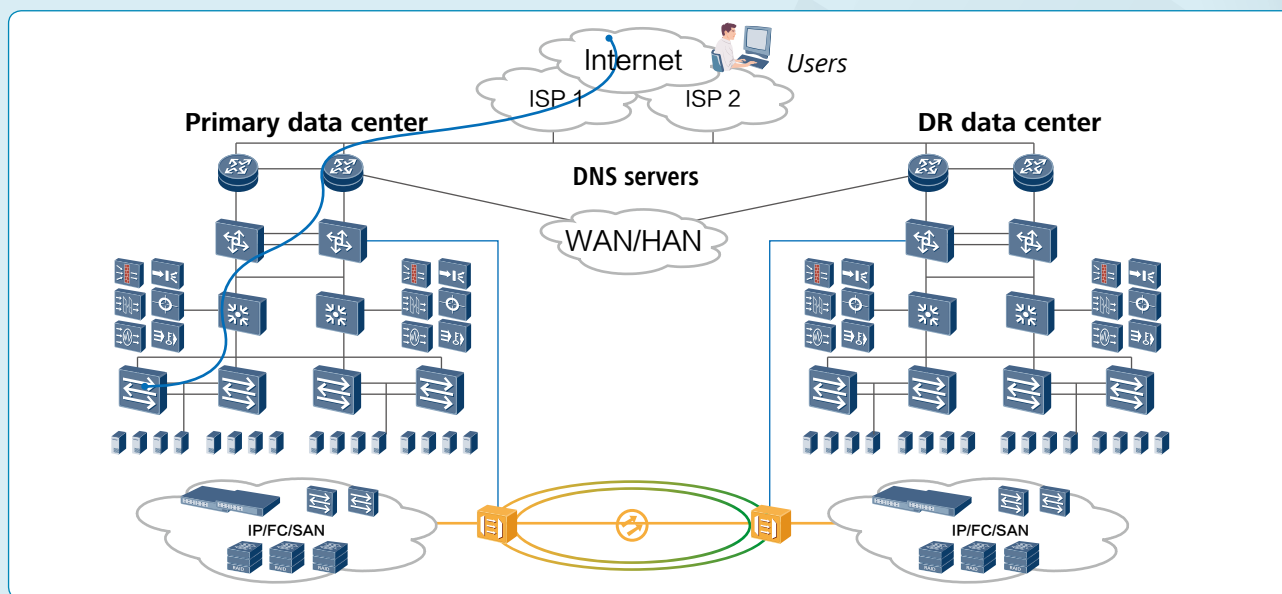


# Solution

In a multi-data-center scenario, a data consistency solution is designed to make all users with the same account access the same piece of physical data (though those users can log in to different data centers). Because all updates occur in the same data center, conflicts caused by asynchronous data replication are prevented. All solutions need to strike a perfect balance among RTO, RPO, and enterprises' organization capabilities.

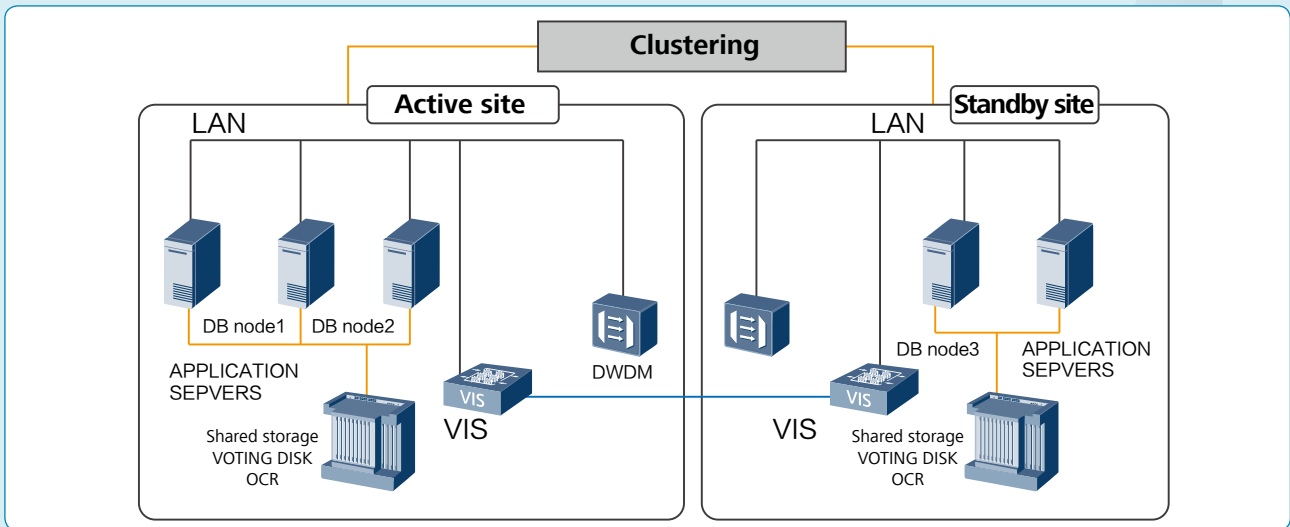
## The solution applies to the following scenarios:

**1.Active and standby data centers:** Because all updates occur in the same database, conflicts caused by data inconsistency are prevented.

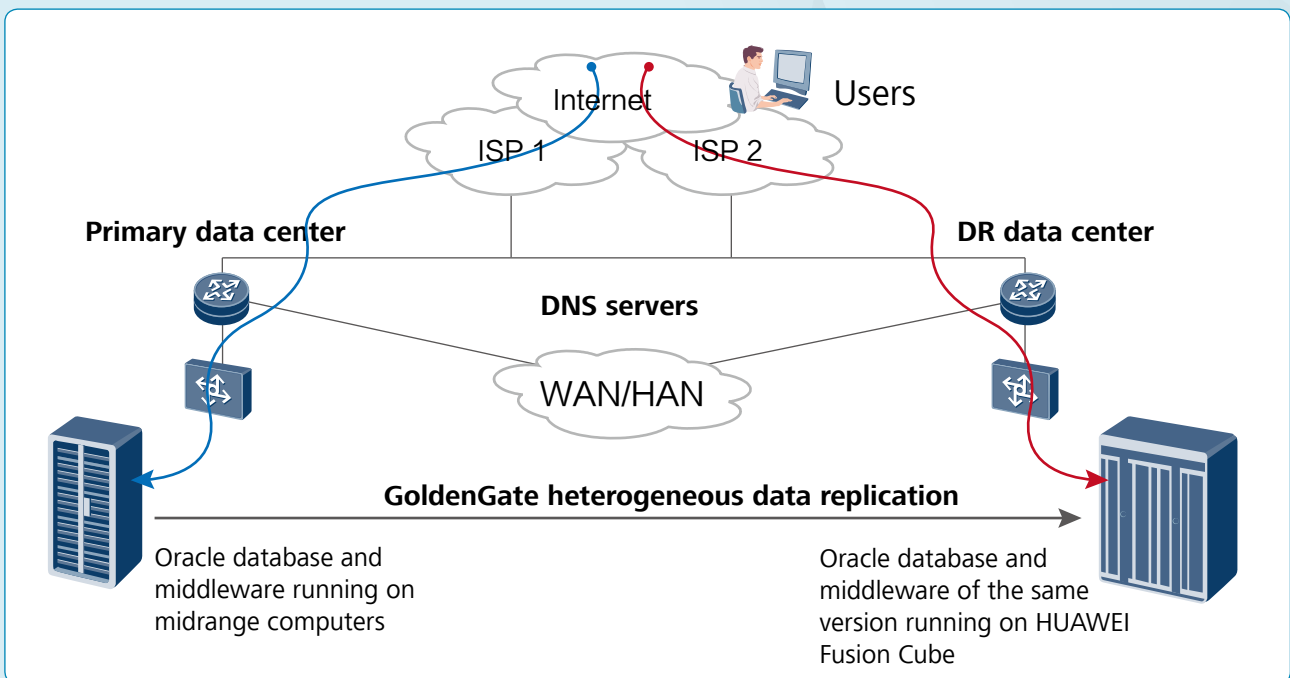


**2.Single database instance deployed in multiple data centers:** In this scenario, data access conflicts do not exist because all database operations take place in all the data centers at the same time. Specifically, instances of a database behave identically, whether they are running in the same or multiple data centers. However, such a solution requires that the distance between data centers be shorter than 40 km.





**3. Multiple databases deployed in multiple data centers:** If the data synchronization latency between databases is not severe, database accounts and application data are synchronously replicated between the databases. In doing so, even though users share their account information in different data centers, their database operations are confined to only one database. If users log in to an "incorrect" data center (where their database does not belong), they can still use their own database. Such a solution works if the distance between active-active data centers is shorter than 1000 km.





# Typical Scenarios

The following sections will elaborate on the process of implementing applicable solutions for the preceding three scenarios.

## 1.Active-Standby Data Centers

### ■ Overview

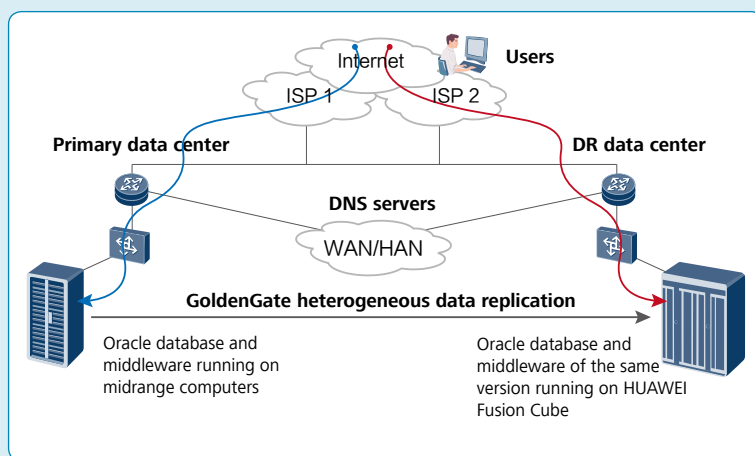
Active-standby data centers are easy to deploy and have been applied in scenarios that are sensitive to business continuity and fault tolerance. However, they are not suitable for scenarios where users are expected to choose a nearby data center for optimal access performance or when service loads need to be balanced between data centers during peak hours. Devices on the primary and DR systems in this scenario can be either homogeneous or heterogeneous.

### ■ Methodology

In active-standby mode, only one data center is active at a time. If the active data center malfunctions, it will be demoted to the standby while the standby data center will be automatically or manually promoted to the active.

### ■ Data Synchronization

The basic way to synchronize data in active-standby mode is to migrate redo logs asynchronously from the active to the standby database. This method has no negative impact on the application layer because it does not change the layer, preventing data loss, and maintaining the high availability of databases. If the access latency between the two databases is not severe and zero data loss is required, data can be synchronously replicated. However, data is usually asynchronously replicated because the impact on the active database can be minimized and the normal operation of the active database is not affected by abnormalities on the network or by the standby database.



### ■ Implementation

The active-standby mode can be homogeneous and heterogeneous. In homogeneous mode, the architectures of the primary and DR systems are the same, while in heterogeneous mode, they are different:

The active-standby mode can be homogeneous and heterogeneous. In homogeneous mode, the architectures of the primary and DR systems are the same, while in heterogeneous mode, they are different:

### For application systems

In this DR solution, FusionCube employs servers, storage systems, and operating system to back up traditional host databases. To ensure application system interoperability, the type of version of database used in the primary center and DR data center must be the same or compatible, so that the primary and standby databases have the same requirements for application systems such as requirements on the database connection mode, session management, transaction control, and SQL statements. In this solution, application systems access the two databases without any difference.

### For databases and middleware

This DR solution can be based on mainstream operating systems including IBM AIX, HP UNIX, and Sun Solaris, and can use common databases and middleware including Oracle 9i, Oracle 10g, Oracle 11g, IBM DB2 9.1, IBM DB2 9.5, IBM DB2 9.7, Sybase ASE 12, Sybase ASE 15, Microsoft SQL Server 2000, Microsoft SQL Server 2005, Microsoft SQL Server 2008, MySQL 5.1, and MySQL 5.5.

### For servers and storage devices

Oracle GoldenGate has the following server, storage device, and network requirements for its primary database:

- CPU

The CPU usage of Oracle GoldenGate is determined by the data load of the primary database. According to pressure tests and real-life deployments, a GoldenGate data replication link usually occupies 2% to 3% of CPU resources, and not more than 7%, with database logs increasing at a speed of 3.5 TB per day.

- Memory

The memory resources assigned to Oracle GoldenGate can be set on the operating system. Memory consumed (usually 40 MB to 55 MB) by a single process for extracting GoldenGate varies depending on the sizes and access frequencies of onsite transaction data in the primary database. When the assigned memory resources are used up, GoldenGate will use hard drives as cache. It is recommended that each GoldenGate data replication link be assigned at least 2 GB of memory.

## Summary

The active-standby mode is easy to implement in technology that brings considerable benefits. It is an ideal entry-level solution for remote DR.



## 2. Single Database Instance Deployed in Multiple Data Centers

### ■ Overview

This method is primarily applicable in scenarios where the infrastructure of a regional data center is insufficient, and, compromising its ability to provide sufficient service quality. Where a minimal system alteration is required, this method proves a valuable reference model.

Only when a disaster is confined in a data center can data reliability be truly ensured.

The method is recommended when:

- Business continuity is valued higher than fault tolerance capability.
- The network infrastructure consists of two or more data centers and network bandwidth between data centers is sufficient.
- No capacity nor desire to modify the application layer.
- A certain period of downtime is acceptable.

### ■ Methodology

When the distance between data centers is less than 40 km, those data centers can function as one logical data center to share the same databases, web server, and load balancer. When the access latency between logical nodes is short enough, components connected to multiple database nodes form a single logical database. The access latency has a direct impact on the overall database performance. To minimize the access latency, a high-speed, low-latency, and high-bandwidth dedicated link is necessary for inter-node communications.

If active-active data centers are deployed to enhance business continuity over fault tolerance, a single database can be deployed in multiple data centers for business continuity while maintaining a certain level of data reliability.

### ■ Data Synchronization

The Oracle Real Application Clusters (RAC) technology can split a RAC database into two or more adjacent data centers. A closer distance between data centers means faster database running, but more adverse impact on business continuity. This is because two data centers located close to each other are more susceptible to damage by tornado, flood, and fire. Besides, as only one database exists in multiple data centers, faults at the database layer may cause all applications to become unavailable. Even though separate and autonomous data centers can isolate faults for higher data reliability, they still lack an efficient means to isolate major faults from spreading to other data centers.

RAC-based DR has no specific requirements for the distance between data centers. The allowed distance range is determined performance, network, storage, and database requirements, as well as network quality between nodes. According to Oracle, a distance greater than 100 km results in low database performance as the long latency causes a split-brain.

### ■ Load Balancing

Since multiple data centers can be regarded as one logical data center in this solution, much of the existing infrastructure is utilized and only a small overhead is incurred due to fault isolation. Traditional seven-layer load balancers can be installed



on web servers in multiple data centers. Domain Name System (DNS)-based applications deployed in each data center for load balancing can temporarily be used to forward access requests during periods of data congestion. Edge load balancers can also be used to less loads between data centers.

Similar to the selection of a database replication/synchronization policy, a load balancing policy must be chosen based on existing infrastructure and the required level of fault tolerance.

### ■ Implementation

#### For application systems

In these scenarios, the connection of application systems is configured based on their network service names rather than their physical IP addresses, because the database has already been taken over by another data center, but the application systems have not. For this reason, application systems must float with the database.

#### For databases and middleware

In these scenarios, databases and middleware must be deployed in clustered mode. Nodes in the cluster must be deployed to different data centers. Besides, databases and middleware clusters must be able to take over services upon disasters and balance service loads during peak hours.

#### For servers and storage devices

In these scenarios, the storage devices must be shared storage that delivers high input/output operations per second (IOPS). They are managed using storage virtualization devices, presenting themselves as a single shared storage product to databases.

In addition, servers in each data center must have the same CPU and operating system models; otherwise, a single database instance cannot run concurrently across multiple data centers.

#### For networks

Such a scenario requires that: Network latency between the data centers is as low as possible; otherwise, the database may be mistakenly taken over by another data center. Efficiency is improved by network acceleration and data exchange speeds are increased by Fibre Channel switches and dense wavelength division multiplexing (DWDM) devices.

## Summary

Because multiple data centers use the same data system, applications and sessions can be taken over within the system, ensuring high reliability, as well as low RTO and RPO. However, this requires a short distance between data centers, minimized network latencies, and powerful local processing capabilities.

## 3 Multiple Databases Deployed in Multiple Data Centers

### ■ Overview

This deployment maximizes the benefits of active-active data centers. Since it allows data centers to be deployed over 1000 km apart, it can deliver high reliability with simple configurations.

### ■ Methodology

The solution allows databases to independently run in different data centers, providing high-performance services for most users, and isolating points of failure to ensure business continuity.

For active-active data centers that are widely spaced apart, it is a technical challenge for multiple users with the same account to log in to those data centers simultaneously. Data synchronization (with latencies usually a few seconds) will cause data conflicts when multiple users update the same pieces of data at the same time unless the application layer is modified in the following way: If a user that attempts to log in to a data center is determined to have already logged in to another data center, that user will receive the connection to the currently logged-in database. To support this mechanism, each data center instance must establish a connection to each data center database. For example, local users in a New York data center have access to local databases and also access to databases in a Boston data center. According to user login locations, data sources can be switched from the New York data center to the Boston data center for optimal access performance, which means that a few HTTP requests are written to databases in the second data center. More jobs need to be done for better data center and database performance.

A DNS server or other dedicated devices can serve as an intelligent load balancer between data centers at the network layer. User requests will then be directed to a nearby data center based on user locations (regions, countries, or continents).

### ■ Data Synchronization

The solution must support bi-directional replication between databases. If the distance between data centers exceeds the upper limit, redo logs received by databases must be read-only. These requirements are vital to data integrity.

### ■ Load Balancing

This solution requires a geographic load balancing algorithm. Specifically, most user requests should be directed to a nearby data center. Such an algorithm can be implemented on a per host or application basis.



### ■ Implementation

#### For application systems

In this solution, both the primary and DR databases should be running all the time. When receiving no external requests, the DR system does not generate database write operations, and can be active all the time and connected to the DR database. In this way, active-active DR is implemented for application systems, further reducing the time and risks during application system switchover.

If the DR database is used to carry the read-only loads of the primary database, dual data sources are recommended for application systems. Read and write data sources are directed to the primary database to enable a quick response to write operations and instant query operations, whereas read-only data sources are directed to the DR database to support the query and report operations that tolerate seconds of latencies.



#### For databases and middleware

For details about Oracle GoldenGate's requirements for the primary database to implement active-active database DR, see "For databases and middleware" in Implementation.

Before the DR database takes over services, all write operations except GoldenGate data replication must be avoided in the DR database. For example, the trigger must be stopped when the database does not come from Oracle or the Oracle database is earlier than Oracle 10.2.0.5 or Oracle 11.2.0.2.

#### For servers and storage devices

For details about Oracle GoldenGate's server, storage device, and network requirements for its primary database, see "For servers and storage devices" in Implementation.

#### For networks

To implement active-active database DR, Oracle GoldenGate has the following requirements:

##### 1 ) For network architecture

The primary data center's network architecture is described as follows:

The core network employs the architecture that integrates the core layer and access layer on a flattened network.

- Based on its functionality, the network architecture is divided into the extranet area, network service area, and service area.
- Compared to the primary data center, the DR data center is capable of a simpler network architecture and uses devices of lower specifications. For example, the Internet egress of the extranet area in the DR data center is not connected by multiple telecom operators. The network architecture varies depending on site requirements.

## 2 ) For data center interconnections

Data center network interconnection is the basis of active-active DR data centers. It is recommended that the primary and DR data centers be connected to each other over a Layer 2 or Layer 3 network. Specifically, a Layer 2 network is used for VM migration and server cluster deployment across data centers, while a Layer 3 network is used for data synchronization between storage devices or databases. If the primary and DR data centers are deployed in the same city, direct optical fiber connections or dedicated cables are recommended for data center interconnection. If they are deployed in different cities, dedicated cables or Multiprotocol Label Switching (MPLS) Virtual Private Network (VPN) connections are recommended.

**Exclusive optical fiber interconnection****Optical:** fibers used to connect between data centers for data exchange and DR backup.

**Advantages:** High bandwidth and low latency because data centers exclusively use the optical fibers for data exchange  
Available to Layer 2 and Layer 3 interconnection

**Disadvantages:** Data center investment is escalated because optical fiber resources need to be established or leased

**Layer 2 VPLS interconnection:** Virtual Private LAN Service (VPLS) is a Layer 2 VPN technology that provides Ethernet-based communication over MPLS networks. The purpose of VPLS is to interconnect multiple Ethernets to the Internet to create a single bridged LAN. By encapsulating a Layer 2 VPN channel using existing Internet or dedicated network resources, VPLS carries data exchanges between data centers, maintains DR backup for business continuity, and is widely applied to interconnected cloud computing data centers.

**Advantages:** No need to create an interconnection plane as a VPN channel is added to the existing network channel to isolate data traffic

**Disadvantages:** 1. Complicated deployment procedures 2. Leased or self-owned MPLS networks required

**Layer 2 EVN interconnection:** Ethernet Virtual Network (EVN) is a brand-new Layer 2 Data Center Interconnect (DCI) technology that provides an overlay of ETH-over-IP. EVN can divide broadcast domains to reduce the risks of broadcast storms.

**Advantages:** Independent of optical fiber and MPLS network resources because Layer 2 EVN interconnection can be flexibly established as long as an IP network is available

Low investment, easy deployment, and simplified O&M

**Disadvantages:** 1. Limited network quality on IP networks 2. Low bandwidth utilization due to overlays

**Layer 3 MPLS VPN interconnection:** By encapsulating a Layer 3 VPN channel using existing Internet or dedicated network resources, MPLS VPN carries data exchanges between data centers, maintains DR backup for business continuity, and is widely applied to interconnected traditional data centers. This connection has similar advantages and disadvantages to Layer 2 VPLS interconnection.



### 3 ) For global traffic management

In an active-active DR data center solution, data centers must collaborate to direct user request to the most appropriate data center, and direct user request to the DR data center when the primary data center crashes. To meet the preceding requirements, the solution must possess the following capabilities:

**Global traffic load balancing:** Active-active application services use the global server load balance (GSLB) technology to serve users at a nearby data center, back up data between data centers in different geographic locations, and balance loads of data centers.

**Active-active network DR requires that:** One GSLB device is deployed in both the primary and DR data centers respectively. Each GSLB device is connected to its egress router in redundant mode to improve reliability.

GSLB devices intelligently resolve domain names for the service IP addresses of application systems.

GSLB devices select the optimal data center for users based on user locations and Internet Service Provider (ISP) links. Specifically, when receiving a DNS request, a GSLB device uses the static or dynamic traffic allocation algorithm (which will be discussed later) to determine which data center is the most appropriate and returns with the IP address of the application service, enabling users to access a nearby data center with the optimal performance and achieving load balancing and redundant backup between data centers.

**Static traffic allocation algorithm:** When a client accesses a GSLB device, the device queries the client's geographic location or carrier information from the pre-set IP address list based on the client's source IP address and uses the queried information to select the IP address of an appropriate data center to respond to the DNS request.



**Dynamic traffic allocation algorithm:** When a client accesses a data center, the two GSLB devices use the dynamic proximity algorithm to detect the client's local DNS server, measure the network speed between the client and each data center, and assign the domain name resolution of the nearest data center to the client based on the test results.

**Health check:** A GSLB device performs health checks for application and database servers, eliminates single points of failure, and enables traffic to bypass a malfunctioning or low-performance server or data center.

A wide range of health check items can be configured to ensure the high availability of application and database servers, including IP address, service, content, and interaction.

**IP address check:** The GSLB device sends a monitoring data packet to the IP address of a node server. If the node server does not respond, the GSLB device will consider it malfunctioning and will not send further service requests to it. Internet Control Message Protocol (ICMP) can be used to detect packages in this check.

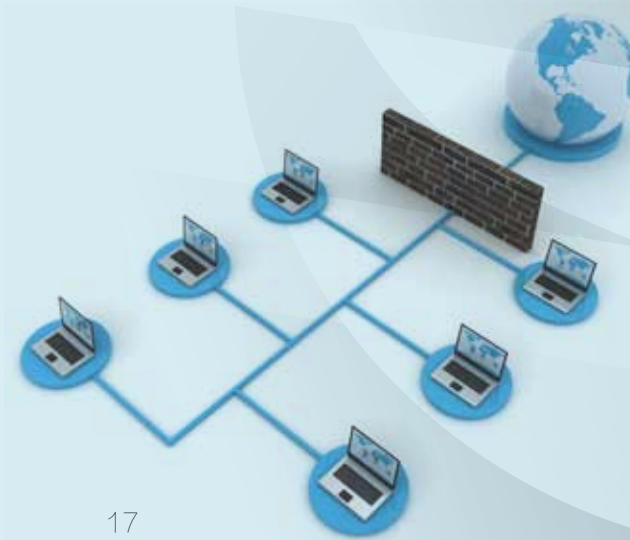
**Service check:** The GSLB device attempts to establish a Transmission Control Protocol (TCP) connection (IP address: service) to a node server. If TCP connection fails, the GSLB device will consider the node server malfunctioning and will not send further service requests to it. TCP can be used to detect packages in this check.

**Content check:** The GSLB device establishes an Open TCP connection (IP address: service) to a node server, sends a request to it, and cuts out the connection after receiving a data package from the node server. If information contained in the package is incorrect, the GSLB device will consider the node server malfunctioning and will not send further service requests to it. HTTP can be used to detect packages in this check.

**Interactive check:** The GSLB device establishes a connection (IP address: service), simulates interactive application sessions, and checks the returned result. If the result is different from the expected, the GSLB device will consider the node server malfunctioning and will not send service requests such as database SQL requests to it afterwards.

## Summary

This solution puts active-active configurations into full play, implements high-level physical isolation for medium-sized areas using simple configurations, and is widely applicable to scenarios that require long-distance transmission but moderate RTO and RPO.



## Solution Summary

---

Whether the evolution towards the long-distance active-active configuration is successful largely depends on the cooperation among individuals in enterprises. Active-active and multi-active high availability solutions not only involve products and devices, but also take end-to-end factors into consideration, meeting customers' service requirements in a wide range of scenarios.

Drawing strength from FusionCube, network products, and Oracle data replication software, Huawei's solutions seamlessly fit into customers' production environments and allow customers to replicate data among heterogeneous storage systems, protecting their initial production investments. The active-active bidirectional DR architecture ensures real-time DR environment monitoring, high resource utilization, and rapid DR switchback. The solutions have the following highlights:

- High-performance DR system, ensuring powerful transaction processing capabilities after service takeover
- High resource utilization, allowing both primary and DR environments to remain active
- Full use of existing devices, protecting original investments
- Secure and reliable DR solution that configures computing and storage nodes in full redundancy and encrypts data during transmission
- Active-active DR solution that enables users to log in to the DR data center in real time and monitor the latest activity status of the DR environment




Copyright © Huawei Technologies Co., Ltd. 2014. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

#### Trademark Notice



**HUAWEI**, and  are trademarks or registered trademarks of Huawei Technologies Co., Ltd.

Other trademarks, product, service and company names mentioned are the property of their respective owners.

#### General Disclaimer

All logos and images displayed in this document are the sole property of their respective copyright holders. No endorsement, partnership, or affiliation is suggested or implied.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.