

Best Practices for Deploying High Availability Architecture on Oracle Cloud Infrastructure

ORACLE REFERENCE ARCHITECTURE | OCTOBER 2018





Revision History

The following revisions have been made to this white paper since its initial publication:

Date	Revision
October 18, 2018	<ul style="list-style-type: none"><li data-bbox="505 548 959 575">• Added information about fault domains.<li data-bbox="505 583 1214 615">• Added a scenario for deployment in a single availability domain.

You can find the most recent versions of the Oracle Cloud Infrastructure white papers at <https://cloud.oracle.com/iaas/technical-resources>.



Table of Contents

Abstract	4
Introduction	4
High Availability Building Blocks	4
Architecting High Availability Solutions	6
Compute High Availability Design	6
Network High Availability Design	9
Load Balancing High Availability Design	9
FastConnect and VPN High Availability Design	12
Storage High Availability Design	18
Database High Availability Design	20
Conclusion	23



Abstract

This reference architecture paper provides Oracle customers and partners with architectural best practices for designing high availability (HA) solutions that will be deployed on Oracle Cloud Infrastructure. It includes detailed descriptions of how to leverage various features and capabilities that are specific to Oracle Cloud Infrastructure.

Introduction

This reference architecture paper provides the best practices needed for planning, designing, and deploying high availability (HA) architectures on Oracle Cloud Infrastructure.

A high availability service or application is one designed for maximum potential uptime and accessibility. To design a high availability architecture, three key elements should be considered—redundancy, monitoring, and failover:

- Redundancy means that multiple components can perform the same task. The problem of a single point of failure is eliminated because redundant components can take over a task performed by a component that has failed.
- Monitoring checks whether or not a component is working properly.
- Failover is the process by which a secondary component becomes primary when a primary component fails.

This paper describes architectural design that focuses on these three key elements. Although high availability can be achieved at many different levels, including the application level and the cloud infrastructure level, this paper focuses on the cloud infrastructure level.

High Availability Building Blocks

An Oracle Cloud Infrastructure *region* is a localized geographic area composed of one or more *availability domains*, each composed of three *fault domains*.

An availability domain is one or more data centers located within a region. Availability domains are isolated from each other, fault tolerant, and unlikely to fail simultaneously. Because availability domains do not share physical infrastructure, such as power or cooling, or the internal availability domain network, a failure that impacts one availability domain is unlikely to impact the availability of others.

A fault domain is a grouping of hardware and infrastructure within an availability domain. Each availability domain contains three fault domains. Fault domains let you distribute your instances so that they are not on the same physical hardware within a single availability domain. As a result, an unexpected hardware failure or a Compute hardware maintenance that affects one fault

domain does not affect instances in other fault domains. You can optionally specify the fault domain for a new instance at launch time, or you can let the system select one for you.

All the availability domains in a region are connected to each other by a low-latency, high-bandwidth network. This predictable, encrypted interconnection between availability domains provides the building blocks for both high availability and disaster recovery.

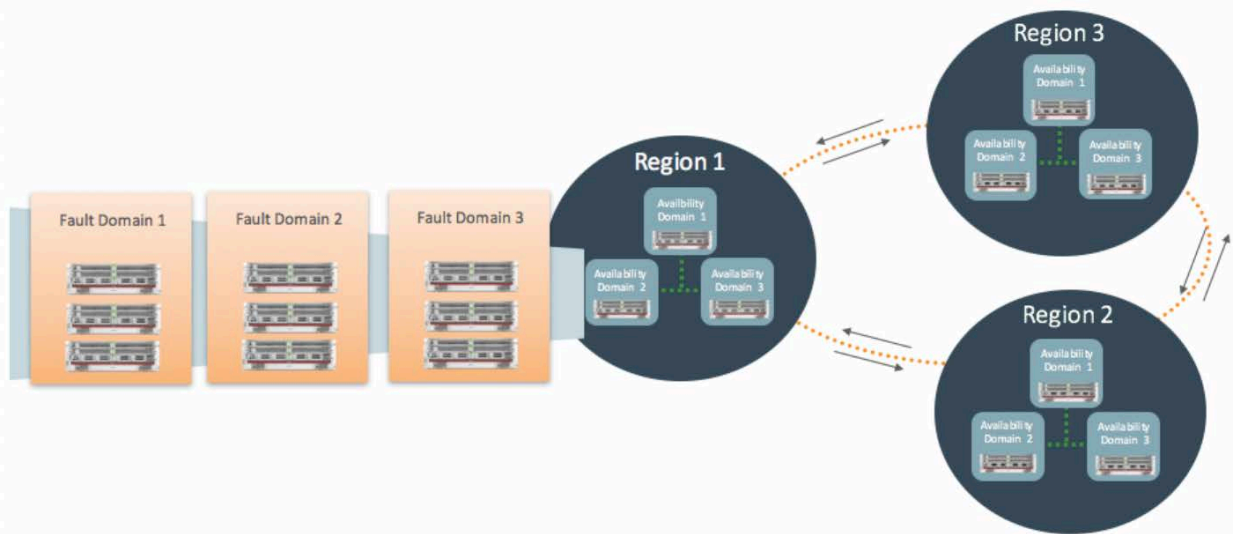


Figure 1. High Availability Building Blocks

Oracle Cloud Infrastructure resources are either specific to a region, such as a virtual cloud network, or specific to an availability domain, such as a Compute instance. When you configure your cloud services, if the services are specific to an availability domain, it is important to leverage multiple availability domains or fault domains to ensure high availability and to protect against resource failure. By creating redundant Compute instances in other availability domains or fault domains, you can avoid an impact to your applications by an issue that affects the primary Compute instance or its domain.

You can design solutions to have multiple regions, multiple availability domains, or multiple fault domains, depending on the class of failures you want protect against.

Architecting High Availability Solutions

This section describes each Oracle Cloud Infrastructure layer and provides detailed best practices and design guidelines for architecting high availability solutions.

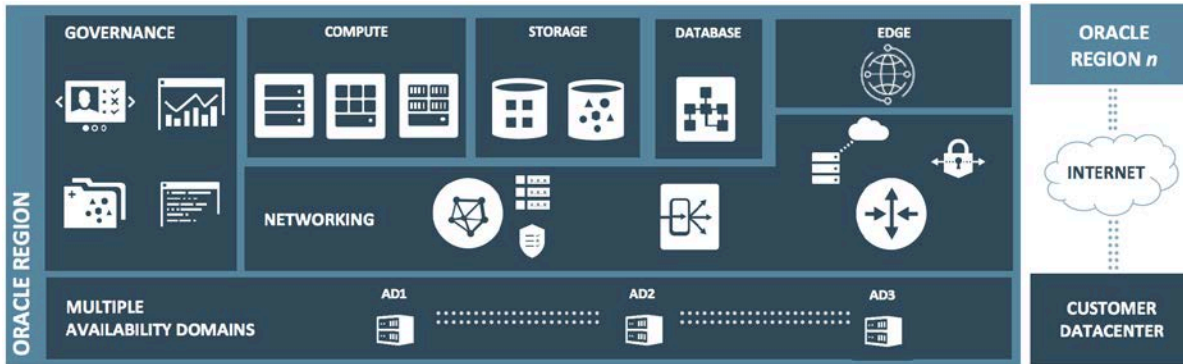


Figure 2. Oracle Cloud Infrastructure High-Level Architecture

Compute High Availability Design

Oracle Cloud Infrastructure Compute provides both bare metal and virtual machine (VM) instances to give you the flexibility to deploy any size server that you need, from a small VM with a single core to a bare metal server with 52 cores and 768 GB of RAM. These options give you the performance, flexibility, and control to run your most demanding applications and workloads in the cloud.

Elimination of a Single Point of Failure

One of the key principles of designing high availability solutions is to avoid single point of failure.

Single Availability Domain Deployment

Each availability domain has three fault domains. By properly leveraging fault domains, you can increase the availability of applications running on Oracle Cloud Infrastructure.

Your application's architecture determines whether you separate or group instances by using fault domains.

- **Scenario 1: Highly Available Application Architecture**

In this scenario, you have a highly available application—for example, two web servers and a clustered database. In this scenario, you group one web server and one database node in one fault domain and the other half of each pair in another fault domain. This architecture ensures that a failure of any one fault domain does not result in an outage for your application.

- **Scenario 2: Single Web Server and Database Instance Architecture**

In this scenario, your application architecture is not highly available—for example, you have one web server and one database instance. In this scenario, both the web server and the database instance must be placed in the same fault domain. This architecture ensures that your application is impacted only by the failure of that single fault domain.

Multiple Availability Domain Deployment

Another approach to high availability is to deploy Compute instances that perform the same tasks in multiple availability domains. This design removes a single point of failure by introducing redundancy.

The following diagram illustrates web server VMs deployed in two availability domains to implement redundancy:

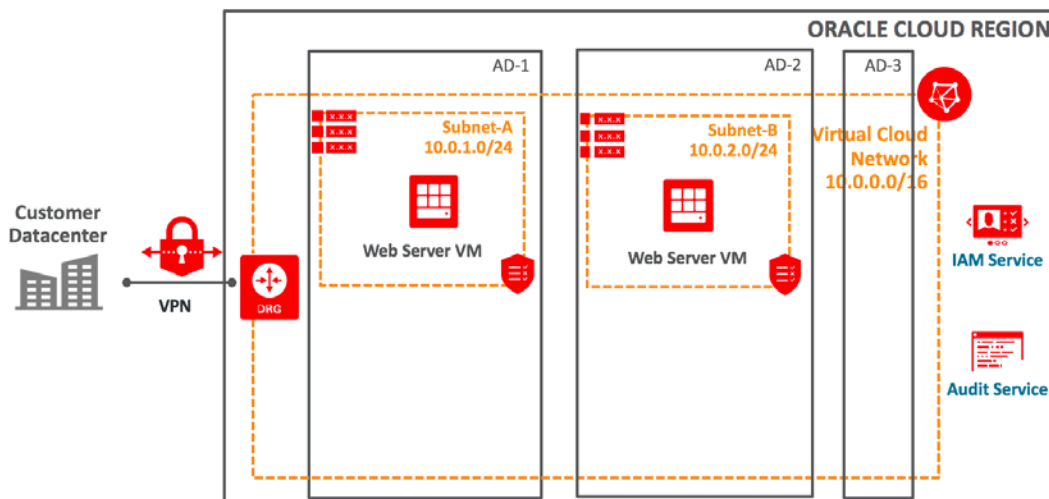


Figure 3. Deploy Web Server VMs in Two Availability Domains

Depending on your system or application requirements, you can implement this architecture redundancy in either standby or active mode:

- In standby mode, a secondary or standby component runs side-by-side with the primary component. When the primary component fails, the standby component takes over. Standby mode is typically used for applications that need to maintain their states.
- In active mode, no components are designated as primary or standby; all components are actively participating in performing the same tasks. When one of the components fails, the related tasks are simply distributed to another component. Active mode is typically used for stateless applications.



Floating IP Addresses

Floating IP addresses of Compute instances, either the secondary private IP address or the reserved public IP address, play a key role in high availability architecture design on Oracle Cloud Infrastructure.

A Compute instance can be assigned a secondary private IP address. If the Compute instance has problems, you can reassign that secondary private IP address to a standby instance in the same subnet to achieve instance failover.

A reserved public IP address can be persistent and exist beyond the lifetime of the Compute instance to which it's currently assigned. In the case of high availability and failover scenarios, you can unassign a reserved public IP address from the primary instance and then reassign it to standby instance.

You can automate this floating IP address failover by leveraging Linux high availability services, such as Corosync or Pacemaker.

Data Availability and Integrity

For a high availability architecture, it's important to ensure that your design protects both the data availability and integrity of your Compute instances. To protect the data availability of your Compute instance, you can either replicate or back up your data to another location.

You can use either synchronous or asynchronous replication to protect your data if your Compute instance fails:

- As described in a previous section, the availability domains of Oracle Cloud Infrastructure are close enough to each other and have a high-performance network to support synchronous replication. If your application needs an instant failover and can't tolerate data loss, we recommend synchronous replication. Because of its network performance requirements, synchronous replication is typically used within one region.
- For applications that need the protection of data availability across regions, we recommend asynchronous replication.

Traditional backups are another way to protect your data. For maximum data durability, don't store your backups in the same availability domain as their original Compute instance. We recommend using Oracle Cloud Infrastructure Object Storage to back up the data of your Compute instance.

For Compute instances with local NVMe drives, a protected RAID array is the most recommended way to protect against an NVMe device failure.



Network High Availability Design

One of the first steps in working with Oracle Cloud Infrastructure is to set up a virtual cloud network (VCN) for your cloud resources. A VCN is a software-defined network that you set up in Oracle Cloud data centers. A subnet is a subdivision of a cloud network. Ensuring the high availability of your network is one of the most important items in your architecture design.

Determining the Right Size of Subnets

Each subnet in a VCN exists in a single availability domain and consists of a contiguous range of IP addresses that do not overlap with other subnets in the cloud network (for example, 172.16.1.0/24). The first two IP addresses and the last one in the subnet's CIDR are reserved by the Oracle Cloud Infrastructure Networking service. You can't change the size of the subnet after it is created, so it's important to think about the size you need before creating subnets. Consider the future growth of your workloads and leave sufficient capacity to meet high availability requirements, such as the need to set up standby Compute instances.

Load Balancing High Availability Design

Oracle Cloud Infrastructure Load Balancing provides automated traffic distribution from one entry point to multiple servers reachable from your virtual cloud network (VCN). The service offers a load balancer with your choice of a public or private IP address and provisioned bandwidth.

The Load Balancing service improves resource utilization, facilitates scaling, and helps ensure high availability. It supports routing incoming requests to various backend sets based on virtual hostname, path route rules, or combination of both.

Public Load Balancer

To accept traffic from the internet, you create a public load balancer. The service assigns it a public IP address that serves as the entry point for incoming traffic. You can associate the public IP address with a friendly DNS name through any DNS vendor.

A public load balancer is regional in scope and requires two subnets, each in a separate availability domain. As a result, a public load balancer is inherently highly available across availability domains. To achieve high availability for your systems, you can put the systems behind a public load balancer. For instance, you can put your web server VMs as backend server sets behind a public load balancer, as illustrated in the following diagram:

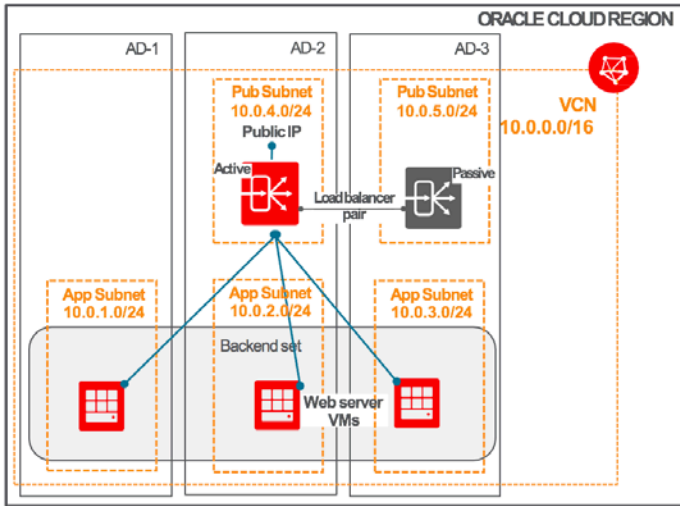


Figure 4. High Availability Architecture with a Public Load Balancer

Private Load Balancer

To isolate your load balancer from the internet and simplify your security posture, you create a private load balancer. The Load Balancing service assigns it a private IP address that serves as the entry point for incoming traffic.

When you create a private load balancer, the service requires only one subnet to host both the primary and standby load balancers. In this case, private load balancer service is bounded within an availability domain.

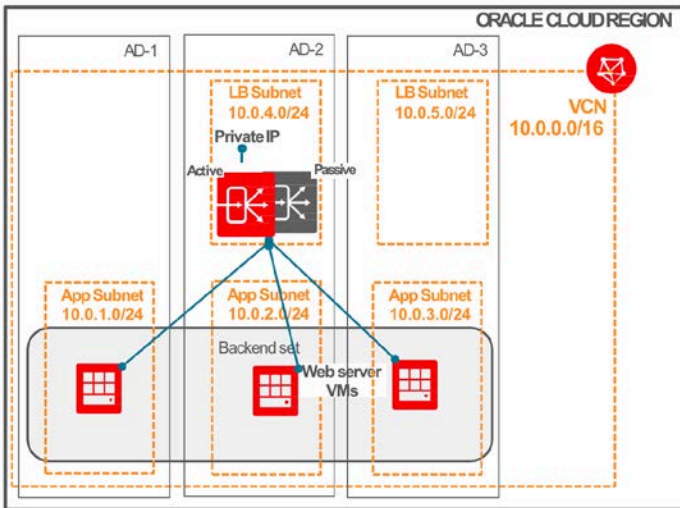


Figure 5. High Availability Architecture with a Private Load Balancer

To provide high availability across availability domains, customers can configure multiple private load balancers on Oracle Cloud Infrastructure and use on-premises or private DNS servers to set up a round-robin DNS configuration with the IP addresses of the private load balancers. You set up this configuration as follows:

1. Deploy two private load balancers, one in each availability domain.
2. Configure two custom DNS VMs in the VCN.
3. Modify the VCN Default DHCP options to use a Custom DNS Resolver and set the DNS servers to the IP addresses of the DNS VMs.
4. Add a new round-robin DNS zone entry for the private load balancer FQDN with a low TTL.
5. Add two A records with the IP addresses of the two private load balancers.
6. Use the FQDN of the private load balancer when accessing the private load balancer.

The following diagram illustrates how to set up high availability private load balancers across available domains:

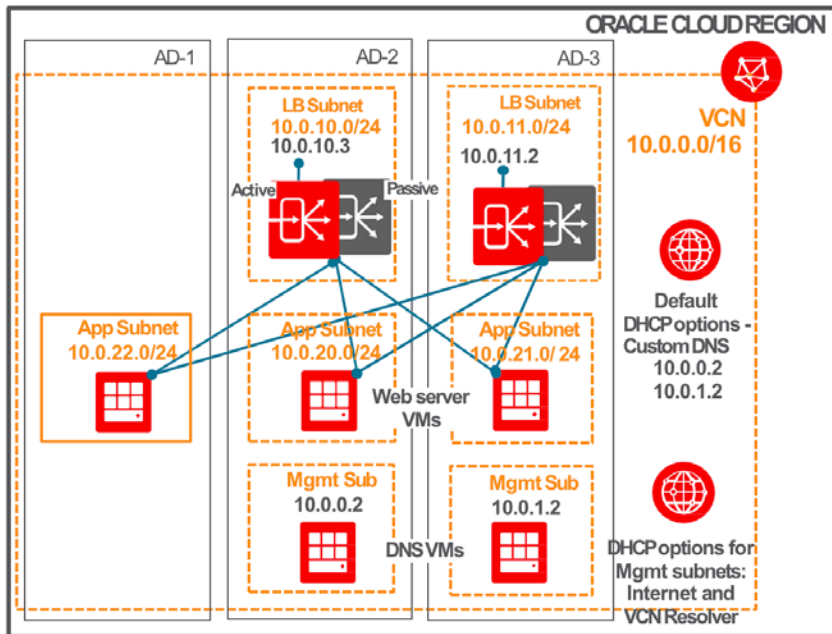


Figure 6. High Availability Architecture with a Private Load Balancer Across Availability Domains



FastConnect and VPN High Availability Design

Highly available, fault-tolerant network connections are key to a well-architected system. This section gives guidelines on how to design your network for redundancy so that it meets the requirements for the Oracle Cloud Infrastructure IPSec VPNs and FastConnect service level agreement (SLA). It discusses high availability options for redundant VPN connections, redundant FastConnect connections, and a FastConnect connection with a backup VPN connection.

An organization's business-availability and application requirements help determine the most appropriate configuration when designing remote connections. Generally, however, you should consider using redundant hardware and network service providers between your location and Oracle's data centers. The most robust option is to use multiple FastConnect connections with circuits from different network service providers.

To achieve high availability for your network, we recommend following best practices:

- Schedule regular maintenance by Oracle, your provider, or your own organization.
- Avoid single points of failure, even if you are planning to use multiple interfaces for availability. High availability connections require redundant hardware, even when connecting from the same physical location.
- Consider a dual provider approach to ensure network diversity when selecting FastConnect providers.
- Provision sufficient network capacity to ensure that the failure of one network connection doesn't overwhelm and degrade redundant connections.

Network High Availability Design with IPSec VPN

You can choose to implement IPSec VPN connections to connect your data center to Oracle Cloud Infrastructure. An IPSec VPN connection is easy to set up and cost-effective.

To enable redundancy, each Oracle Cloud Infrastructure dynamic routing gateway (DRG) has multiple VPN endpoints so that each IPSec VPN connection consists of multiple redundant IPSec tunnels that use static routes to route traffic. To ensure high availability, you must set up VPN connection availability within your internal network to use either path when needed as illustrated in following diagram:

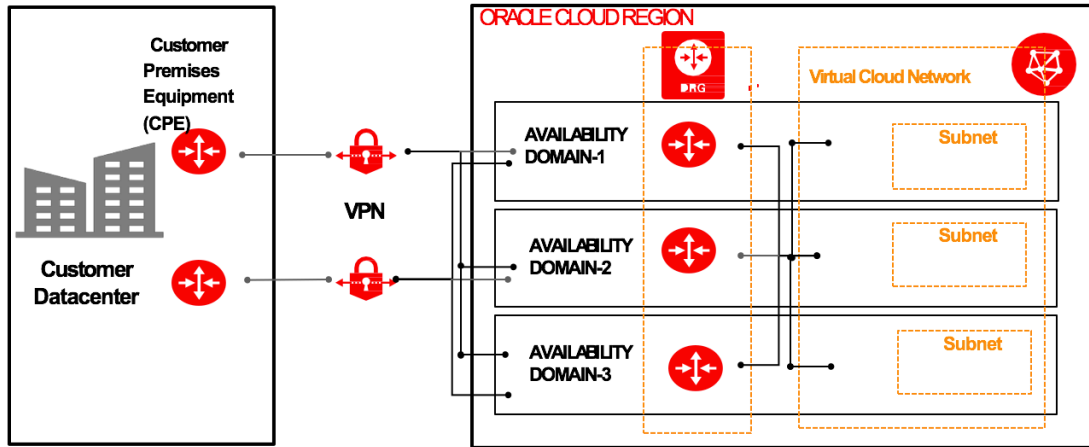


Figure 7. IPsec VPN High Availability Design

Using a Broad CIDR as a Static Route

If your data centers span multiple geographical locations, we recommend using a broad CIDR (0.0.0.0/0) as a static route in addition to the CIDR of the specific geographical location. This broad CIDR provides high availability and flexibility to your network design.

For instance, the following diagram shows two networks in separate geographical areas that each connect to Oracle Cloud Infrastructure. Each area has a single on-premises router, so two IPsec VPN connections can be created. Note that each IPsec VPN connection has two static routes: one for the CIDR of the particular geographical area, and a broad 0.0.0.0/0 static route.

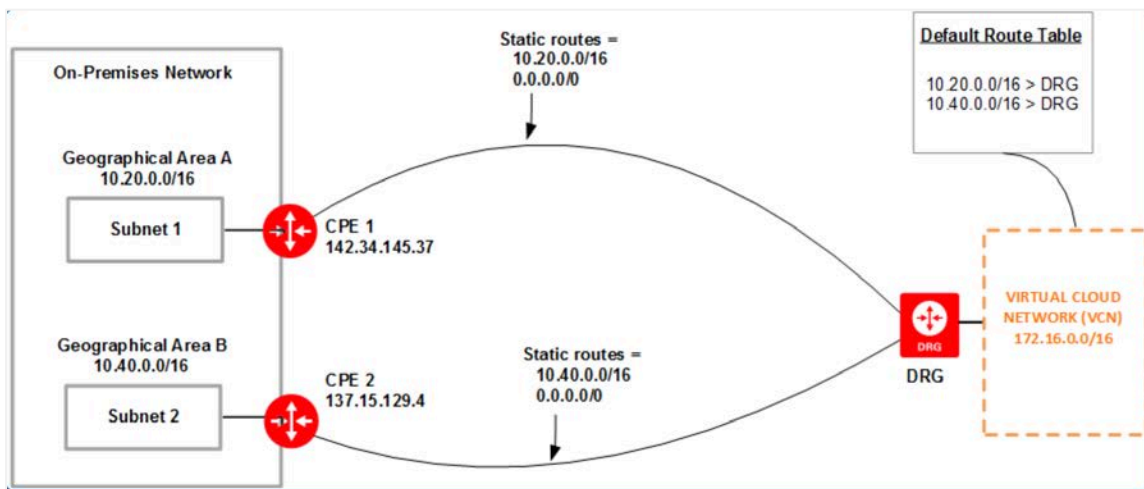


Figure 8. IPsec VPN with Multiple Geographical Locations

In one scenario, the CPE 1 router in the preceding diagram goes down. If Subnet 1 and Subnet 2 can communicate with each other, the VCN is still able to access the systems in Subnet 1 because of the 0.0.0.0/0 static route that goes to CPE 2. The following diagram illustrates this scenario:

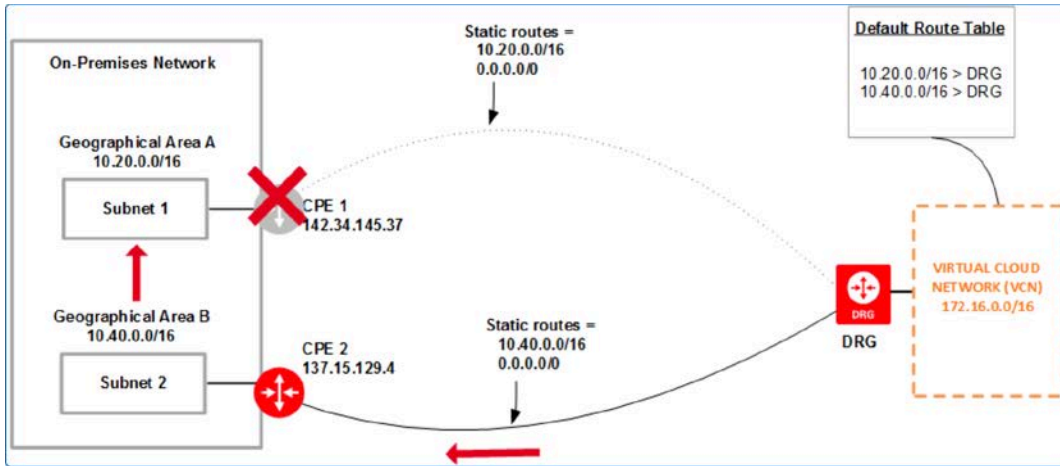


Figure 9. IPsec VPN High Availability Scenario 1

In another scenario, you add a new geographical area with Subnet 3 and connect it to Subnet 2. You would add a route rule to your VCN's route table for Subnet 3 so that the VCN can reach systems in Subnet 3 without creating a new VPN connection because of the 0.0.0.0/0 static route that goes to CPE 2. The following diagram illustrates this scenario:

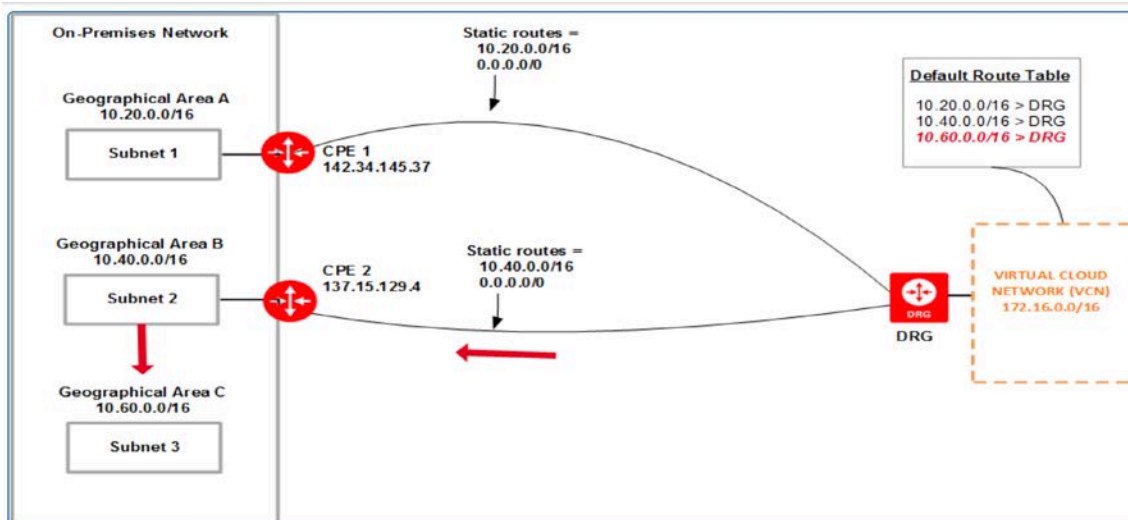


Figure 10. IPsec VPN High Availability Scenario 2



Network High Availability Design with FastConnect

Oracle Cloud Infrastructure FastConnect provides an easy way to create a dedicated, private connection between your data center and Oracle Cloud Infrastructure. FastConnect provides higher-bandwidth options and a more reliable and consistent networking experience compared to internet-based connections.

With FastConnect, you can choose to use private peering, public peering, or both.

- Use private peering to extend your existing infrastructure into a virtual cloud network (VCN) in Oracle Cloud Infrastructure (for example, to implement a hybrid cloud, or in a lift-and-shift scenario). Communication across the connection is with IPv4 private addresses (typically RFC 1918).
- Use public peering to access public services in Oracle Cloud Infrastructure without using the internet (for example, to access Object Storage, the Oracle Cloud Infrastructure Console and APIs, or public load balancers in your VCN). Communication across the connection is with IPv4 public IP addresses. Without FastConnect, the traffic destined for public IP addresses would be routed over the internet. With FastConnect, that traffic goes over your private physical connection.

You can either connect directly to Oracle Cloud Infrastructure routers in provider points-of-presence (POPs) or use one of Oracle's many partners to connect from POPs around the world to their Oracle Cloud Infrastructure Networking resources. Oracle provides features that allow you to build fault-tolerant connections, including multiple POPs per region and multiple FastConnect routers per POP.


FastConnect Redundancy

To avoid a single point of failure with redundancy, Oracle Cloud Infrastructure provides the following features:

- Multiple FastConnect locations within each metro area
- Multiple routers in each FastConnect location
- Multiple physical circuits in each FastConnect location

Oracle handles the redundancy of the routers and physical circuits in the FastConnect locations. In your network design with FastConnect, we recommend considering the following redundancy configurations for your high availability requirements:

- **Availability domain redundancy:** Connect to any FastConnect location and access services located in any availability domain within a region. This configuration provides availability domain resiliency via multiple POPs per region. Peering connections terminate on routers in the POP.

- 
- **Data center location redundancy:** Connect at two different FastConnect locations per region.
 - **Router redundancy:** Connect to two different routers per FastConnect location.
 - **Circuit redundancy:** Have multiple physical connections at any of the FastConnect locations. Each of these circuits can have multiple physical links in an aggregated interface/LAG, which adds another level of redundancy.
 - **Partner/provider redundancy:** Connect to the FastConnect locations by using single or multiple partners.

Based on the location of the on-premises data center, you can establish a FastConnect connection in one of the following ways:

- **Colocation** (port speed of 10 Gbps): By colocating with Oracle in a FastConnect location
- **Oracle provider** (port speeds in 1-Gbps and 10-Gbps increments): By connecting to an Oracle provider

In a colocation scenario, a *cross-connect* is the physical cable connecting your existing network to Oracle in the FastConnect location. When you are provisioning your FastConnect service, we recommend that you set up at least two cross-connects. Each cross-connect should connect to a different router, so that a failure in one router does not impact your connection to Oracle Cloud Infrastructure resources. After making the first cross-connect, you can request that the second one be provisioned on a different Oracle FastConnect router than the first one. You should provision new virtual circuits on both redundant links, which ensures connectivity between your on-premises network and Oracle Cloud Infrastructure VCNs if one router fails.

For the Oracle provider scenario, we recommend that you set up redundant circuits with two different FastConnect locations by the same provider or different providers. With this configuration, you can have redundancy on both the circuits and the data center levels. The following diagram illustrates FastConnect connection with two virtual circuits and two different FastConnect locations:

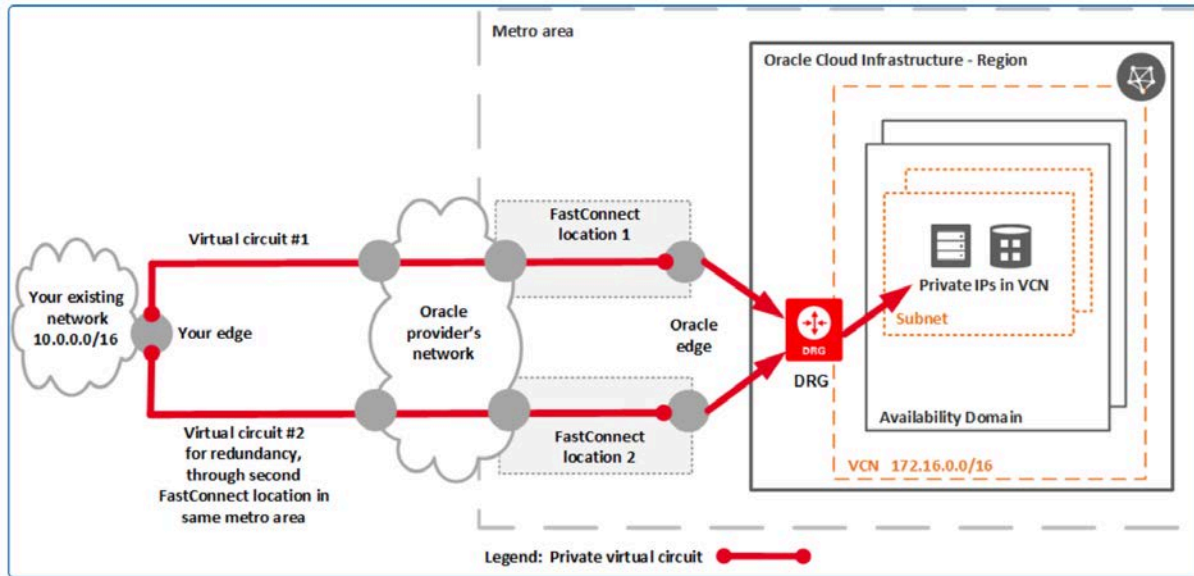


Figure 11. FastConnect High Availability Design

Oracle's FastConnect partners have redundant links to the Oracle network. As a customer of the partner, you should have redundant links to the partner's network. These connections should be on different routers, both in your network and in the partner's network. When you provision virtual circuits, provision them across your multiple provider links.

Avoiding Impact During Planned Maintenance

When you want to perform maintenance on one of your routers, you can configure your Border Gateway Protocol (BGP) local preference on routes learned over their virtual circuit so that the local preference is higher on the router that will stay in service. BGP local preference is used to modify outbound traffic preference in an on-premises network.

You can modify traffic from Oracle to your network by using BGP AS prepending. On the router where the maintenance will be performed, prepend your local BGP AS number. Doing so causes the Oracle Cloud network to prefer the FastConnect virtual circuit that has the shorter AS path.

After you modify the BGP local preference and AS prepending, monitor your router's virtual circuit interface counters and verify that the in and out packet counter rates are very low. The only traffic remaining on the link should be BGP protocol traffic.

Continuous Testing of Redundant Paths

During normal operation, we recommend using all available paths between your on-premises network and the Oracle Cloud. Doing so ensures that if a failure occurs, your redundant path is already working. Alternatively, using an active/backup design means that you trust that your

backup path will work during a failure. For this reason, you should consider using equal BGP local preference and BGP AS path length.

Using Both IPsec VPN and FastConnect

To have an additional level of redundancy, you can set up both IPsec VPN and FastConnect to connect your on-premises data centers to Oracle Cloud Infrastructure. When you set up both an IPsec VPN connection and FastConnect virtual circuits to the same DRG, remember that the IPsec VPN uses static routes but FastConnect uses BGP. Oracle Cloud Infrastructure advertises a route for each of your VCN's subnets over the FastConnect virtual circuit BGP session, and overrides the default route selection behavior to prefer BGP routes over static routes if a static route overlaps with a route advertised by your on-premises network. The following diagram illustrates this configuration:

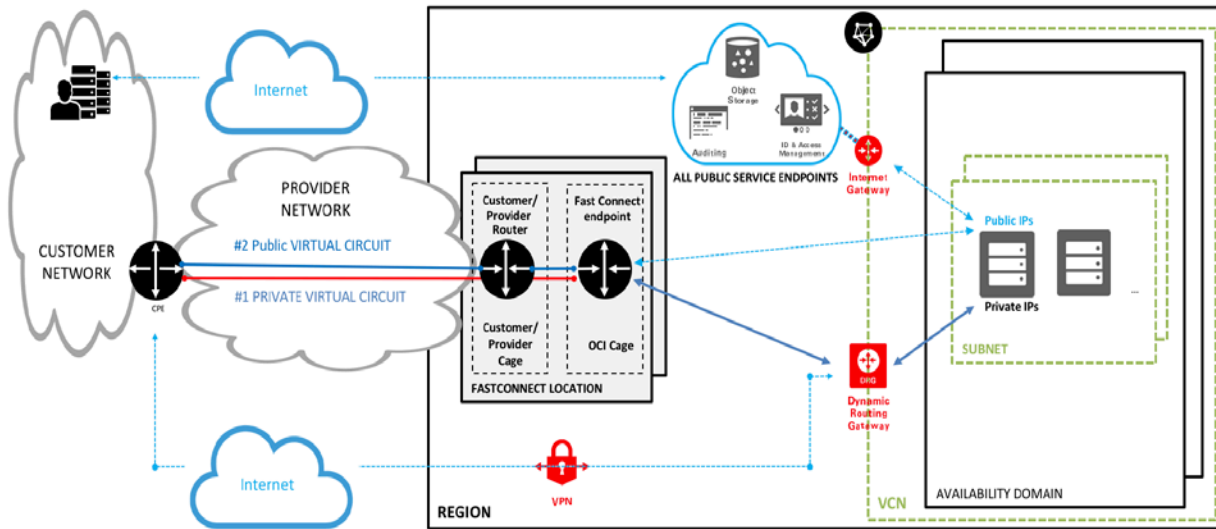



Figure 12. Use Both IPsec VPN and FastConnect

Storage High Availability Design

Oracle Cloud Infrastructure provides the following storage services:

- Block Volume
- Object Storage
- File Storage

Oracle Cloud Infrastructure Block Volume enables you to dynamically provision and manage block storage volumes. You can create, attach, connect, and move volumes as needed to meet your storage and application requirements. When a volume is attached and connected to an instance,




you can use it like a regular hard drive. Volumes can also be disconnected and attached to another Compute instance while the data on the volume is maintained.

Oracle Cloud Infrastructure Object Storage is an internet-scale, high-performance storage platform that offers reliable and cost-efficient data durability. The Object Storage service can store an unlimited amount of unstructured data of any content type, including analytic data and rich content, like images and videos. Object Storage is a regional service and is available across all the availability domains within a region. Data is stored redundantly across multiple storage servers and across multiple availability domains.

Oracle Cloud Infrastructure File Storage provides a durable, scalable, distributed, enterprise-grade network file system. You can connect to a File Storage file system from any bare metal, virtual machine, or container instance in your Virtual Cloud Network (VCN). You can also access a file system from outside the VCN by using Oracle Cloud Infrastructure FastConnect and IPsec VPNs. Large Compute clusters of thousands of instances can use File Storage for high-performance shared storage, and it provides redundant storage for resilient data protection.

To achieve high availability and durability, we recommend the following best practices for the storage layer:

- Use Object Storage to back up application data. Data is stored redundantly across multiple storage servers across multiple availability domains. Data integrity is actively monitored by using checksums, and corrupt data is detected and automatically repaired. Any loss in data redundancy is automatically detected and corrected, without any customer impact.
- Use Block Volume policy-based backups to perform automatic, scheduled backups and retain them based on a backup policy. Consistently backing up your data allows you to adhere to your data compliance and regulatory requirements.
- If you need an immediate, point-in-time, direct disk-to-disk copy of your block volume, use the Block Volume cloning feature. Volume cloning is different than snapshots because there is no copy-on-write or dependency to the source volume. No backup is involved. The clone operation is immediate, and the cloned volume becomes available for use right after the clone operation is initiated. You can attach and use the cloned volume as a regular volume as soon as its state changes to available.
- If you need to safeguard data against accidental or malicious modifications by an untested or untrusted application, use a block volume with a read-only attachment. A read-only attachment marks a volume as read-only, so the data in the volume is not mutable. You can also use read-only attachments when you have multiple Compute instances that access the same volume for read-only purposes. For example, the instances might be running a web front end that serves static product catalog information to clients.

- 
- When your workload requires highly available shared storage with file semantics, and you need built-in encryption and snapshots for data protection, use File Storage. File Storage uses the industry-standard Network File System (NFS) file access protocol and can be accessed concurrently by thousands of Compute instances. File Storage can provide high-performance and resilient data protection for your applications. The File Storage service runs locally within one availability domain. Within an availability domain, File Storage uses synchronous replication and high availability failover to keep your data safe and available.
 - If your application needs high availability across multiple availability domains, use GlusterFS on top of the Block Volume service.
 - Plan and size your storage capacity by considering future growth needs.

Database High Availability Design

The Oracle Cloud Infrastructure Database service lets you quickly launch an Oracle Database System (DB System) and create one or more databases on it. The Database service supports several types of DB Systems, ranging in size, price, and performance.

Using Exadata DB Systems

Exadata DB Systems allow you to leverage the power of Exadata within the Oracle Cloud Infrastructure. An Exadata DB System consists of a quarter rack, half rack, or full rack of Compute nodes and storage servers, tied together by a high-speed, low-latency InfiniBand network and intelligent Exadata software. You can configure automatic backups, optimize for different workloads, and scale up the system to meet increased demands.

Exadata DB systems provide built-in high availability capabilities. All the existing best practices with your on-premises Exadata DB systems are applicable.

Using 2-node RAC DB Systems

Oracle Cloud Infrastructure offers 2-node RAC DB Systems on virtual machine Compute instances. Because 2-node RAC DB systems provide built-in high-availability capabilities, we recommend using 2-node RAC DB Systems for your solutions that require high availability.

You can configure the Database service to automatically back up to Oracle Cloud Infrastructure Object Storage.

The following diagram shows the deployment of a 2-node RAC DB System to support the high availability of a two-tier web application:

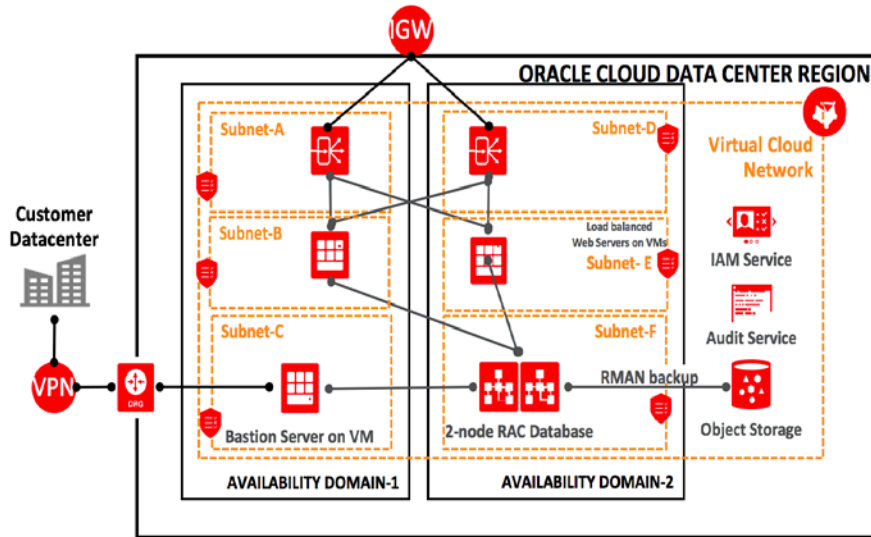


Figure 13. A 2-Node RAC DB System Supports the High Availability of a Two-Tier Web Application

Working with Data Guard

For solutions with a single-node DB system, we recommend using Oracle Data Guard to achieve high availability. Data Guard ensures high availability, data protection, and disaster recovery for enterprise data.

Implementation of Data Guard in the Oracle Cloud Infrastructure Database service requires two databases, one in a primary role and one in a standby role. The two databases compose a Data Guard association. Most of your applications access the primary database. The standby database is a transactional consistent copy of the primary database.

To improve availability and disaster recovery, we recommend placing the DB System of the standby database in a different availability domain from the DB System of the primary database. The high-performance network between Oracle Cloud Infrastructure availability domains enables this deployment.

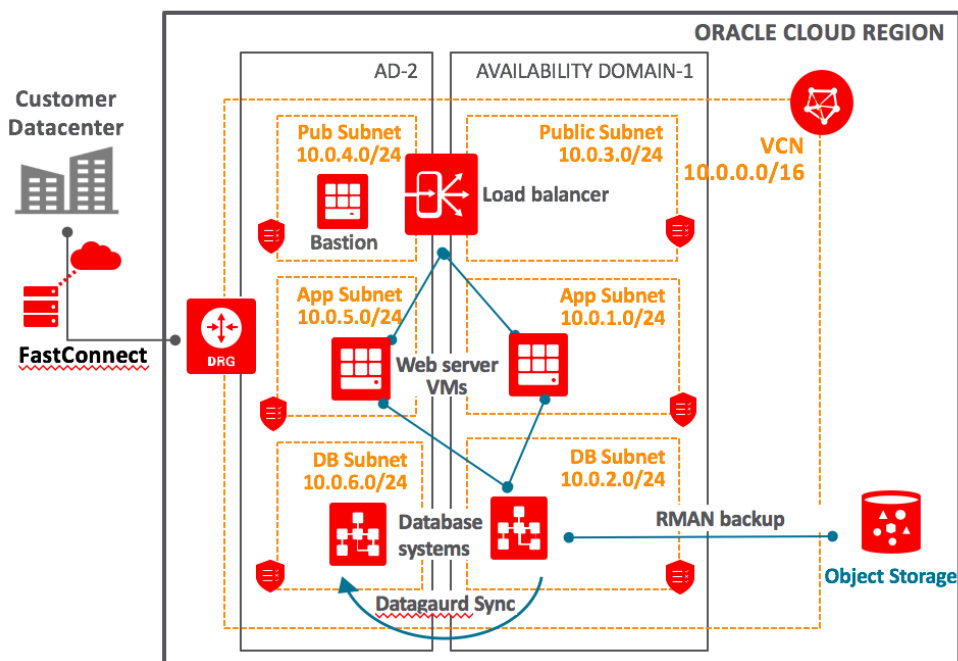


Figure 14. Using Data Guard for a High Availability Database Design


Data Guard maintains the standby database by transmitting and applying redo data from the primary database. If the primary database becomes unavailable, you can use Data Guard to switch the standby database to the primary role.

You can perform following actions with Data Guard configuration to support high availability:

- **Switchover:** Reverses the primary and standby database roles. Each database continues to participate in the Data Guard association in its new role. A switchover ensures no data loss. You can use a switchover before you perform planned maintenance on the primary database.
- **Failover:** Transitions the standby database into the primary role after the existing primary database fails or becomes unreachable. A failover might result in some data loss when you use Maximum Performance protection mode.
- **Reinstate:** Reinstates a database into the standby role in a Data Guard association. You can use the `reinstate` command to return a failed database to service after correcting the cause of the failure.

Automated CPU and Storage Scaling

To achieve high availability for your solutions, you must ensure that your DB Systems have sufficient capacity. Database services on Oracle Cloud Infrastructure can dynamically scale CPU cores or database storage based on the different shapes of your Database service.



For DB Systems based on bare metal Compute instances, we recommend that you start with minimum CPU cores and dynamically increase the number of CPU cores as needed.

For DB Systems based on virtual machine Compute instance, you can dynamically increase the storage size.

Conclusion

When planning any deployment, planning for availability is a key concern. This reference architecture paper provides guidance to help design high availability (HA) solutions on Oracle Cloud Infrastructure, including the compute, network, storage, and database layers, and provides several best practices to help guide your planning:

- Eliminate single points of failure with redundancy.
- Deploy your application or solution components across multiple regions, multiple availability domains within a region, or multiple fault domains within an availability domain.
- Design with future growth in mind and ensure that you have sufficient resource capacity.
- Leverage Oracle Database Data Guard and dynamic scaling capabilities.
- Replicate your application data across availability domains.
- Automate problem resolution and failover processes.

Oracle Cloud Infrastructure is continuously evolving with new features. You can find the latest information, documents, and training at <https://cloud.oracle.com>.




Oracle Corporation, World Headquarters

500 Oracle Parkway
Redwood Shores, CA 94065, USA

Worldwide Inquiries

Phone: +1.650.506.7000
Fax: +1.650.506.7200

CONNECT WITH US

 blogs.oracle.com/oracle

 facebook.com/oracle

 twitter.com/oracle

 oracle.com

Integrated Cloud Applications & Platform Services

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. 1018

Best Practices for Deploying High Availability Architecture on Oracle Cloud Infrastructure

October 2018

Authors: Changbin Gong and Adeel Amin



Oracle is committed to developing practices and products that help protect the environment.