This guide provides an overview of the options, recommendations, and general concepts that you need to know before you deploy a high-availability (HA) SAP NetWeaver system on Google Cloud.

This guide assumes that you already have an understanding of the concepts and practices that are generally required to implement an SAP NetWeaver high-availability system. Therefore, the guide focuses primarily on what you need to know to implement such a system on Google Cloud.

If you need to know more about the general concepts and practices that are required to implement an SAP NetWeaver HA system, see:

- The SAP best practices document <u>Building High Availability for SAP NetWeaver and SAP HANA on Linux</u> (https://support.sap.com/content/dam/SAAP/SAP_Activate/AGS_70.pdf)

- <u>The SAP NetWeaver documentation</u> (https://help.sap.com/viewer/product/SAP_NETWEAVER/ALL/en-US)

This planning guide focuses solely on HA for SAP NetWeaver and does not cover HA for database systems. For information about HA for SAP HANA, see the <u>SAP HANA high availability and disaster recovery planning guide</u> (/solutions/sap/docs/sap-hana-hadr-planning-guide).

The following diagram shows a basic Linux HA cluster that uses Pacemaker cluster software.

The cluster includes two hosts: a primary host and a secondary host. Each host is located in a different zone within the same region.

The active Central Services instance and the inactive Enqueue Replication Server (ERS) instance are on the primary host. The active ERS instance and the inactive Central Services instance are on the secondary host. Each Central Services and ERS pair has its own virtual IP address (VIP). In the diagram, "Central Services" represents either ABAP SAP Central Services or, for a Java stack, SAP Central Services.
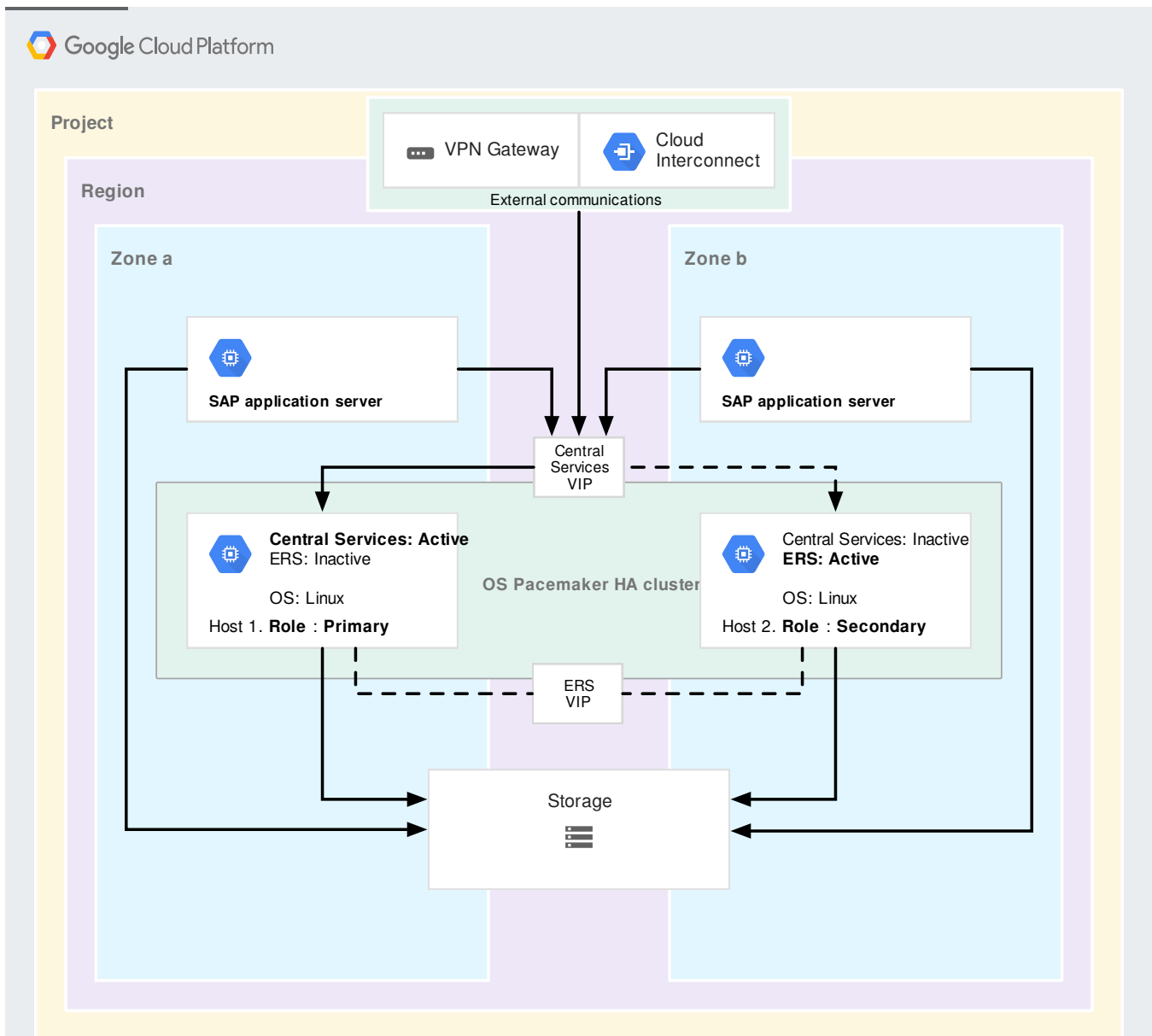
Google Cloud is highly available by design, with a redundant infrastructure of data centers around the world that contain zones designed to be independent from each other. Zones usually have power, cooling, networking, and control planes that are isolated from other zones. If a single failure event were to occur, in most cases, it would affect only a single zone.

In some cases, you might meet your availability requirements without implementing all of the traditional, on-premises safeguards against hardware, storage, and networking failures, which could save you both time and money.

Before you design and implement your high-availability strategy on Google Cloud, review the Google Cloud Service Level Agreements (/terms/sla/).

For general information about the reliability, privacy, and security of Google Cloud, see Trusted Infrastructure (https://support.google.com/googlecloud/answer/6056635?hl=en).

You define a high-availability (HA) cluster for SAP NetWeaver on Google Cloud by using the same types of third-party HA cluster software that you might use in an on-premises installation. The HA cluster software monitors the health of the systems and manages the failover when problems occur.

You can use a number of different HA cluster software solutions, such as the following:

- Red Hat Enterprise Linux (RHEL) for SAP Solutions

- SUSE Linux Enterprise Server (SLES) for SAP Applications

- Windows Server Failover Clustering

Recent versions of both RHEL and SLES include integrated HA support that is enabled specifically for Google Cloud. To check if your Linux version includes Google Cloud-enabled HA support, look for "GCP-HA" in the table in Operating system support for SAP NetWeaver on Google Cloud (/solutions/sap/docs/netweaver-os-support#quick_reference_table).

On Windows Server you use Windows Server Failover Clustering (WSFC) to create the HA cluster, as described in Running Windows Server Failover Clustering (/compute/docs/tutorials/running-windows-server-failover-clustering).

On Google Cloud, the routing of incoming traffic to the active node in a WSFC cluster is managed by Cloud Load Balancing, which does not require an alias IP or static route VIP implementation.

Cloud Load Balancing uses health checks to determine the active node.

Deploy the nodes of your HA cluster across two or more Compute Engine zones within the same region. Deploying the nodes in different zones ensures that they are on different physical machines and also protects against the very unlikely possibility of a zonal failure.

Keeping the zones within the same region ensures that the nodes are close enough geographically to meet SAP latency requirements for high-availability systems.

To support high availability, Compute Engine VMs are backed by live migration and automatic restart.

Compute Engine monitors the state of the underlying infrastructure. When an infrastructure maintenance event occurs, Compute Engine automatically migrates your instance away from the event and, if possible, keeps your instance running during the migration. No user intervention is required.

In the case of major outages, there might be a slight delay between when the instance goes down and when it is available.

In most cases, live migration events occur without impacting the HA cluster. However, test your HA cluster by simulating a live migration of the active host after your HA cluster is set up and the systems are running, especially if your HA cluster monitor is configured with a low failover threshold. For more information about simulating a live migration event, see Testing your availability policies (/compute/docs/instances/setting-instance-scheduling-options#testingpolicies).

A migrated instance is identical to the original instance, including the instance ID, private IP address, and all instance metadata and storage.

By default, standard instances are set to live migrate. We recommend not changing this setting.

For more information, see Live migrate (/compute/docs/instances/setting-instance-scheduling-options#live_migrate).

If your instance is set to terminate when there is a maintenance event, or if your instance crashes because of an underlying hardware issue, you can set up Compute Engine to automatically restart the instance. By default, instances are set to automatically restart. We recommend not changing this setting.

For more information about automatic restart, see Automatic restart (/compute/docs/instances/setting-instance-scheduling-options#autorestart).

The SAP NetWeaver global file system is a single point of failure that needs to be available to all of the SAP NetWeaver instances in your HA system. To ensure the availability of the global file system on Google Cloud, you can use either highly available shared storage or replicated zonal persistent disks.

For a high-availability shared storage solution you can use third-party file-sharing solutions, such as Elastifile (https://www.elastifile.com/) (acquired by Google Cloud) or NetApp Cloud Volumes (https://www.netapp.com/us/cloud-marketplace/google-cloud-platform.aspx). Google Cloud provides an NFS file server solution, Filestore (/filestore/docs/), but Filestore does not currently provide a file server that is highly available across zones.

For replication of zonal persistent disks for Linux systems, you can use a Distributed Replicated Block Device (DRDB) to replicate the persistent disks that contain the SAP global file system between the nodes in your HA cluster.

Although Compute Engine regional persistent disks offer synchronously replicated block storage across zones, they are not currently supported for SAP NetWeaver HA systems.

For more information about storage options on Google Cloud, see:

- File Servers on Compute Engine (/solutions/filers-on-compute-engine)

- Compute Engine storage options (/compute/docs/disks/)

When you set up your network for your HA cluster, in addition to completing the steps in Creating a network (/solutions/sap/docs/netweaver-deployment-linux-dm#creating_a_network), you need to complete the following HA-specific tasks:

- Choose your VIP implementation for Linux systems, as described in the following section. Windows systems use an internal load balancer, which doesn't require the same VIP solutions as Linux systems.

- Define the communication path between the SAP Central Services instance and the Enqueue Replication Server instance.

- Define firewall rules to support your defined communication paths.

To reroute traffic in Linux environments when a failover occurs, you define a VIP, which can be switched back and forth between the VMs in the HA cluster.

On Google Cloud, VIPs are implemented slightly differently than they are in on-premises installations, in that when a failover occurs, gratuitous ARP requests cannot be used to announce the change. Instead, you can implement a VIP address for an SAP NetWeaver HA cluster by using either a Google Cloud static <u>route</u> (/vpc/docs/routes) or a Google Cloud <u>alias IP</u> (/vpc/docs/alias-ip) address.

The static route implementation is recommended for multi-zone SAP NetWeaver HA clusters. Multi-zone deployments help ensure that a single-zone failure does not take down your SAP NetWeaver system.

However, if your network architecture does not support static routes or you need to avoid solutions like IP overlay or avoid complex routing, you can use the alias IP address implementation with a single-zone SAP NetWeaver HA cluster. Alias IP addresses are not recommended for multi-zone deployments because the reallocation of the alias might not be guaranteed in the event of a zonal failure.

Depending on whether you choose the static route implementation or the alias IP implementation, the requirements for the IP address that you use for your VIP are different.

The use of a static route requires that you select an IP address for your VIP that meets the following criteria:

- Outside of the IP ranges of your existing Virtual Private Cloud subnets where the VMs reside.

- Does not conflict with any external IP addresses in your extended network.

Consult with your network administrator to determine a suitable IP address for a static route implementation.

To implement your VIP with an alias IP address, you use an IP address from the IP range of your subnetwork. For more information about VIPs on Google Cloud, see Best Practices for Floating IP Addresses (/solutions/best-practices-floating-ip-addresses).