1. Arquitetura Proposta da Aplicação em Nuvem

1.1. Contextualização

Este documento apresenta a arquitetura de uma aplicação em nuvem projetada para operar com alta disponibilidade, resiliência e escalabilidade. A solução foi desenvolvida considerando os requisitos de aplicações modernas de comércio eletrônico, garantindo continuidade do serviço, tolerância a falhas de zona e capacidade de suportar aumentos repentinos de demanda.

A arquitetura faz uso dos serviços gerenciados do **Google Cloud Platform (GCP)**, aproveitando suas funcionalidades nativas de balanceamento global, instâncias gerenciadas, redes privadas, banco de dados como serviço (PaaS) e mecanismos automáticos de failover.

1.2. Objetivos da Arquitetura

A solução visa alcançar:

- Alta disponibilidade (**24/7**)
- Resiliência a falhas de zona
- Escalabilidade automática sob demanda
- Segurança reforçada por IAM e firewall
- Armazenamento confiável e gerenciado em banco de dados PaaS
- Fault tolerance e recuperação automatizada em todas as camadas

1.3. Visão Geral da Arquitetura

A aplicação está hospedada na região **us-central1**, organizada em três zonas de disponibilidade:

- us-central1-a
- us-central1-b
- us-central1-c

A comunicação é estruturada dentro de uma **VPC Network** dividida em:

- **subnet-app** → máquinas virtuais e aplicação
- **subnet-data** → banco de dados (Cloud SQL) e componentes relacionados

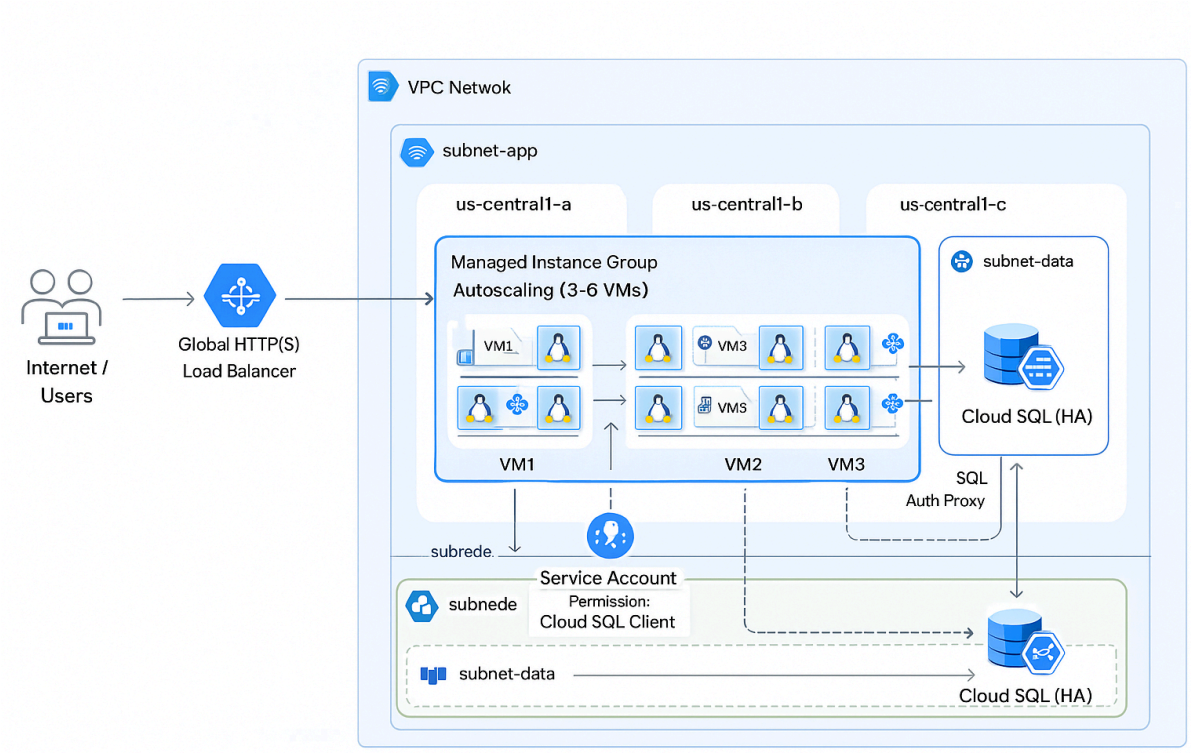
O tráfego externo é recebido por um **Global HTTP(S) Load Balancer**, que faz a distribuição inteligente para o backend.

As instâncias de aplicação são executadas em um **Managed Instance Group (MIG) regional**, configurado com:

- 3 a 6 instâncias Linux
- Autoscaling baseado em métricas (CPU e demanda)
- Distribuição entre múltiplas zonas
- Health checks contínuos

O banco de dados utiliza **Cloud SQL em modo HA**, com réplicas de failover automáticas e backups habilitados.

1.4. Diagrama Arquitetural



2. Descrição da Arquitetura da Solução em Nuvem

Esta seção detalha formalmente os componentes que integram a solução, abordando máquinas virtuais, balanceadores, banco de dados, segurança e mecanismos de failover.

2.1. Topologia da Rede (VPC Network)

A solução está implantada em uma **VPC Network própria**, garantindo isolamento lógico entre recursos. A VPC é segmentada em duas sub-redes:

- **subnet-app** – destinada às máquinas virtuais e à aplicação web.
- **subnet-data** – destinada ao banco de dados gerenciado (Cloud SQL).

Ambas as sub-redes operam exclusivamente com endereçamento interno, e a comunicação entre elas é limitada por regras de firewall.

2.2. Camada de Entrada (Load Balancer)

O tráfego externo chega ao ambiente através de um **Global HTTP(S) Load Balancer**, responsável por:

- Recepção e terminação de tráfego HTTPS
- Roteamento das requisições para as VMs saudáveis
- Monitoramento contínuo via **health checks**
- Failover automático caso instâncias ou zonas falhem

O Load Balancer é um serviço global que opera no edge da Google Cloud, garantindo baixa latência e alta disponibilidade.

2.3. Camada de Aplicação (Managed Instance Group)

A aplicação é executada em um **Managed Instance Group (MIG)** configurado com:

- Instâncias Linux criadas a partir de **Instance Template**
- Distribuição regional entre:
  - `us-central1-a`
  - `us-central1-b`
  - `us-central1-c`

Escalabilidade Automática

O MIG possui política de autoscaling configurada:

- **Mínimo:** 3 VMs
- **Máximo:** 6 VMs
- Baseado em:
  - Uso de CPU
  - Carga de requisições

    - Health checks

O escalonamento garante elasticidade da aplicação, que aumenta ou reduz o número de VMs conforme demanda.

Segurança das VMs

As VMs implementam as seguintes medidas:

- Regras de firewall restringindo acesso apenas ao Load Balancer
- SSH bloqueado para acesso público (habilitado apenas para IPs específicos)
- Atualizações controladas via Instance Template
- Comunicação com banco por rede privada

2.4. Camada de Dados (Cloud SQL)

O banco de dados utiliza o serviço **Google Cloud SQL**, configurado como:

- **Modo Regional (HA)** – instância primária + réplica de failover
- **Private Service Connect** – acesso exclusivo via VPC
- **Backups automáticos diários**
- **Replicação automática entre zonas**

Acesso Seguro ao Banco

As VMs acessam o Cloud SQL usando:

- **Service Account exclusiva**
- Permissão IAM: `roles/cloudsql.client`
- **Cloud SQL Auth Proxy**, garantindo:
    - Conexões criptografadas
    - Autenticação sem senhas

    - Redução de superfícies de ataque

2.5. Mecanismos de Failover

A arquitetura incorpora failover em todas as camadas:

Failover do Load Balancer

- Remove automaticamente instâncias não saudáveis do backend.
- Redireciona tráfego para zonas saudáveis.

Failover do Managed Instance Group

- Recriação automática de VMs falhas.
- Redistribuição de instâncias em zonas saudáveis.

Failover do Banco de Dados

- Cloud SQL HA realiza failover automático para a réplica.
- Conexões são restabelecidas via Private Service Connect sem intervenção humana.