

Descripción de la Arquitectura y Justificación

La solución propuesta se basa en una arquitectura distribuida en AWS diseñada para soportar cargas variables, alta disponibilidad y escalabilidad automática. Desde el punto de vista del usuario final, la experiencia se optimiza mediante la entrega de contenido estático (HTML, CSS, JavaScript e imágenes) a través de CloudFront, la red de entrega de contenido (CDN) de AWS, lo que reduce la latencia al servir los recursos desde ubicaciones geográficas cercanas.

El tráfico de usuarios se canaliza a través de un Application Load Balancer (ALB), que distribuye de manera inteligente las peticiones entre dos grupos de instancias en un clúster multi-AZ: uno dedicado al frontend (interfaz en JavaScript) y otro al backend (lógica de negocio). Esta distribución garantiza que, en caso de fallos o picos de tráfico, la infraestructura se ajuste automáticamente mediante Auto Scaling, manteniendo la continuidad del servicio.

En la capa de backend, se implementa un enfoque dual en la persistencia de datos: se utiliza Amazon RDS para manejar datos estructurados y transaccionales, y DynamoDB para operaciones de baja latencia sobre datos semi-estructurados. Además, el backend se integra con dos microservicios externos que aportan funcionalidades adicionales sin comprometer la independencia y modularidad de la aplicación.

Esta arquitectura no solo ofrece resiliencia y escalabilidad mediante la distribución de la carga y la implementación en múltiples zonas de disponibilidad, sino que también facilita la actualización y el mantenimiento de cada componente de forma independiente. La separación clara entre frontend y backend, junto con la utilización de servicios gestionados, permite optimizar los tiempos de respuesta y reducir la complejidad operativa, adaptándose a las necesidades del negocio en tiempo real.

