



SISTEMAS DISTRIBUTIVOS

CONCEPTO DE ESCALABILIDAD

CONCEPTO DE ESCALABILIDAD

La escalabilidad en sistemas distribuidos se refiere a la capacidad de un sistema para incrementar su rendimiento y mantener un funcionamiento eficiente a medida que aumenta la carga de trabajo o el número de usuarios. Este concepto se convierte en un pilar en el diseño de arquitecturas distribuidas, pues permite que el sistema crezca de forma ordenada sin comprometer la calidad del servicio (Boronat Seguí, 2012).

Escalabilidad vertical y horizontal

Existen dos enfoques principales:

- **Escalabilidad vertical.** Consiste en aumentar la capacidad de un único nodo, agregando más recursos como CPU, memoria o almacenamiento.
 - 🔗 **Ejemplo.** Migrar una base de datos a un servidor con mayor cantidad de núcleos y memoria RAM para procesar más consultas simultáneas.
- **Escalabilidad horizontal.** Implica añadir más nodos al sistema, permitiendo que las solicitudes se distribuyan entre múltiples servidores.
 - 🔗 **Ejemplo.** Utilizar un clúster de servidores web en un balanceador de carga para manejar un mayor tráfico de usuarios en una plataforma de e-learning.

Importancia de la escalabilidad

En sistemas distribuidos, la escalabilidad garantiza:

- Que el sistema pueda responder a un crecimiento en la cantidad de usuarios.
- Que las operaciones mantengan tiempos de respuesta aceptables.
- Que los costos de operación se ajusten de forma proporcional al crecimiento.

Un sistema no escalable se vuelve ineficiente y puede colapsar ante picos de demanda, generando pérdidas económicas y de confianza para los usuarios (Muñoz Escoí, 2013).

Ejemplos prácticos de escalabilidad

- **Netflix.** Utiliza escalabilidad horizontal en su infraestructura de microservicios, permitiendo desplegar múltiples instancias de servicios de transmisión de video en diferentes regiones para atender picos de demanda.
- **Google Search.** Distribuye las solicitudes de búsqueda entre miles de servidores, garantizando tiempos de respuesta rápidos sin importar el volumen de consultas.
- **Tiendas en línea durante eventos de alta demanda (Black Friday).** Utilizan arquitecturas escalables en la nube para agregar instancias de servidores temporalmente y soportar grandes cantidades de transacciones.

Factores que facilitan la escalabilidad

Para que un sistema sea escalable, se considera:

- **Diseño modular.** Separar funcionalidades para que puedan escalar de forma independiente.
- **Uso de bases de datos distribuidas.** Permite que los datos se consulten y se procesen en paralelo.
- **Balanceadores de carga.** Distribuyen las solicitudes de manera equitativa entre los nodos disponibles.
- **Automatización.** Permite crear y destruir instancias, según la demanda en arquitecturas basadas en la nube.

La escalabilidad en sistemas distribuidos, es la capacidad del sistema para adaptarse de forma flexible al crecimiento de usuarios y de datos sin sacrificar el rendimiento ni la disponibilidad. Se logra mediante estrategias de escalabilidad vertical y horizontal, balanceo de carga, bases de datos distribuidas y una arquitectura modular. Gracias a la escalabilidad, los sistemas pueden evolucionar con las necesidades del negocio, garantizando confiabilidad y eficiencia incluso en escenarios de alta demanda (Muñoz Escó, 2013).