



entendiendo el teorema CAP

... y una breve introducción a los sistemas distribuidos





... una breve introducción

Vivimos en un mundo cada vez más interconectado, donde las aplicaciones y los servicios suelen depender de múltiples sistemas distribuidos. Entender el Teorema CAP nos permite diseñar estos sistemas de una manera más eficiente y confiable





Problemáticas comunes

1

Cómo puedo hacer que mi aplicación esté disponible siempre que el usuario necesite hacerlo

2

Cómo puedo hacer que mis usuarios tengan una buena experiencia?

3

Cómo puedo hacer que los datos reflejados sean los reales, no corrompidos o inexistentes



Definición de un Sistema distribuido

Un sistema distribuido es un grupo de computadoras independientes que interactúan como un solo sistema coherente



- Escalabilidad
- Redundancia
- Compartir recursos
- Rendimiento mejorado
- Tolerancia a fallos
- Compartir recursos de manera eficiente



Principales desafíos

- Latencia y ancho de banda
- Fallos y recuperaciones
- Seguridad
- Consistencia de datos






Por qué es difícil?

- **Latencia variante:** En sistemas distribuidos, no se puede asumir que la comunicación es instantánea. La latencia entre nodos puede variar, lo que complica la sincronización y la coherencia.
- **Estados Globales:** No hay una noción simple de "tiempo global" o "estado global" que todas las computadoras en el sistema compartan
- **Recuperación de fallos parciales:** La recuperación de fallos parciales es complicada. No siempre es fácil determinar si un nodo ha fallado o simplemente está tardando mucho en responder
- **Transiciones y Consistencia:** Mantener la consistencia de las transacciones en múltiples nodos es un desafío



... aca se aparece el teorema CAP



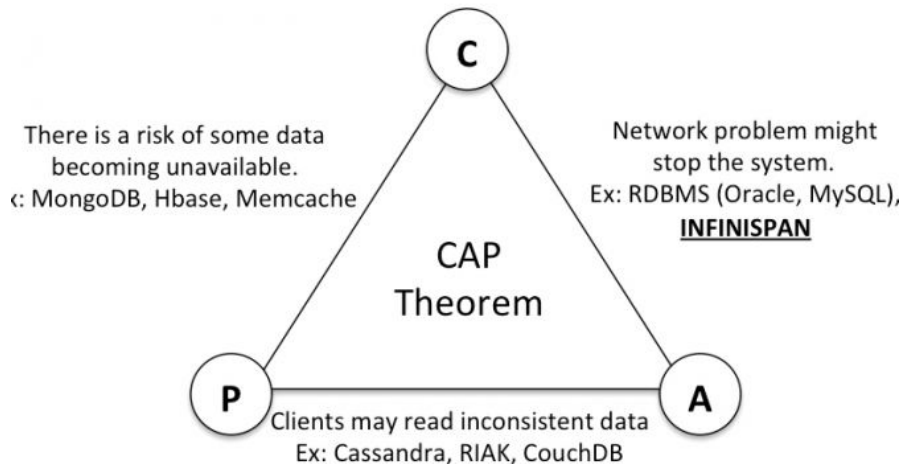
“En un sistema distribuido, no se pueden garantizar simultáneamente la Consistencia, la Disponibilidad y la Tolerancia a la Partición (CAP). Se puede optimizar solo para dos de estos atributos a la vez”.

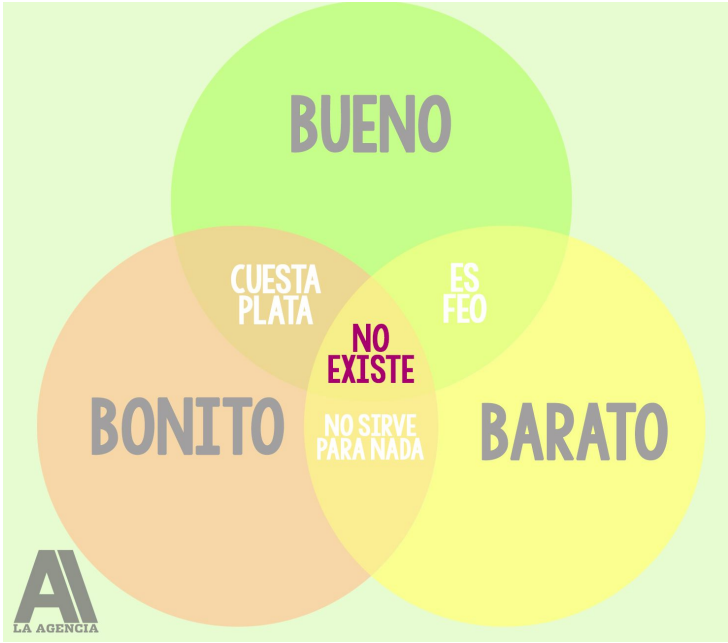
El teorema CAP también se llama Teorema de Brewer, porque fue propuesto por primera vez por el profesor Eric A. Brewer durante una charla que dio sobre computación distribuida en 2000. Dos años más tarde, los profesores de MIT, Seth Gilbert y Nancy Lynch, publicaron una prueba de la "Conjetura de Brewer".

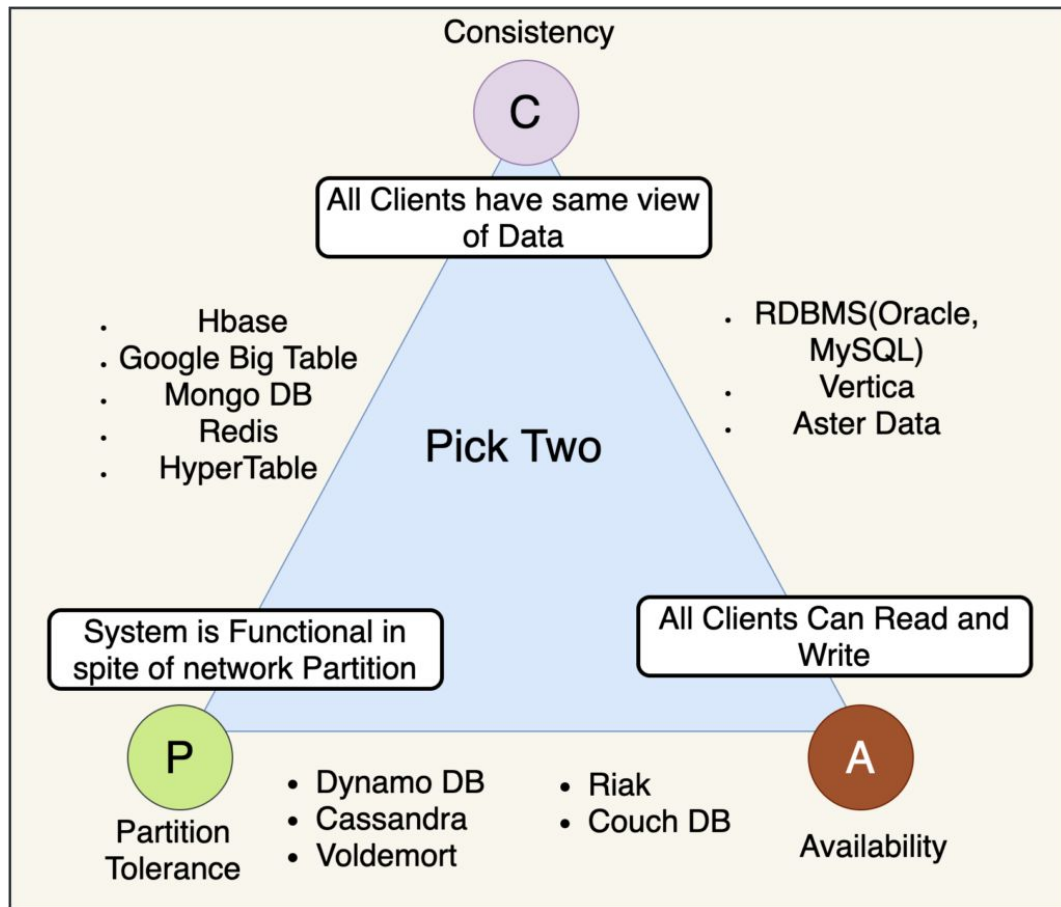
¿que es el teorema CAP?

¿Alguna vez ha visto un anuncio de un paisajista, pintor de casas o algún otro comerciante que comience con el titular "*Barato, rápido y bueno: elija dos*" ?

El teorema CAP aplica un tipo similar de lógica a los sistemas distribuidos, es decir, que un sistema distribuido puede ofrecer solo dos de las tres características deseadas: *consistencia*, *disponibilidad*, y *tolerancia de partición* (el 'C', 'A' y 'P' en CAP).

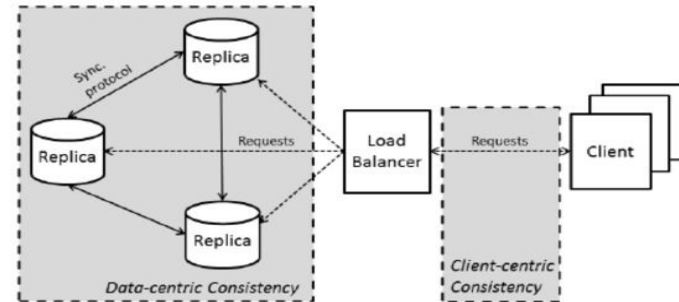






Consistencia

Consistencia significa que todos los clientes ven los mismos datos al mismo tiempo, independientemente del nodo al que se conecten. Para que esto suceda, siempre que se escriban datos en un nodo, se debe reenviar o replicar al instante a todos los demás nodos del sistema antes de que la escritura se considere 'satisfactoria'.





Disponibilidad

Disponibilidad significa que cualquier cliente que realiza una solicitud de datos obtiene una respuesta, incluso si uno o más nodos están inactivos. Otra forma de indicar esto: todos los nodos activos del sistema distribuido devuelven una respuesta válida para cualquier solicitud, sin excepción.





Tolerancia a Particiones

Una partición es un quiebre de las comunicaciones dentro de un sistema distribuido: una conexión perdida o temporalmente retardada entre dos nodos. La tolerancia de las particiones significa que el clúster debe continuar trabajando a pesar de las interrupciones de comunicación que se produzcan entre los nodos del sistema.





Combinacion CA (Consistencia y Disponibilidad)

Sistemas CA se enfocan en mantener una visión uniforme y actualizada de los datos (consistencia) mientras aseguran que todas las solicitudes obtienen una respuesta (disponibilidad).

- **Ventajas:** Velocidad de acceso y manipulación de datos, idóneos para entornos de red fiables.
- **Desventajas:** Poca o nula tolerancia a fallos de red y particiones, lo que los hace vulnerables en entornos distribuidos.
- **Ejemplos:** Las bases de datos monolíticas tradicionales son el ejemplo más cercano, aunque son raros en aplicaciones distribuidas modernas.



Combinacion CP (Consistencia y Tolerancia a Particiones)

Sistemas CP aseguran que todos los nodos en una red vean los mismos datos (consistencia) y pueden funcionar incluso cuando algunos nodos no pueden comunicarse (tolerancia a la partición).

- **Ventajas:** Manejo robusto de errores y datos consistentes a lo largo del tiempo.
- **Desventajas:** Sacrificio en disponibilidad; las solicitudes pueden sufrir retrasos si hay una partición en la red.
- **Ejemplos:** HBase, MongoDB y otros almacenes de datos orientados a consistencia son representativos de esta categoría



Combinacion AP (Disponibilidad y Tolerancia a Particiones)

Sistemas AP priorizan estar siempre en línea (disponibilidad) y funcionar a pesar de interrupciones en la comunicación entre nodos (tolerancia a la partición).

- **Ventajas:** Altamente resilientes, ideales para aplicaciones donde el tiempo de inactividad es inaceptable.
- **Desventajas:** Las operaciones pueden resultar en estados inconsistentes temporalmente hasta que las particiones se resuelvan.
- **Ejemplos:** Couchbase, Cassandra y Riak son ejemplos de bases de datos diseñadas con disponibilidad y tolerancia a la partición en mente



Por qué no los tres?!



Realidad

La realidad del teorema CAP emerge de la naturaleza de las redes y la información. Aquí hay razones clave por las cuales no podemos tener Consistencia, Disponibilidad y Tolerancia a la Partición al mismo tiempo

Latencia en la red

En un mundo ideal sin latencia, podríamos pensar en alcanzar los tres. Sin embargo, en el mundo real, la latencia es inevitable. Transmitir información toma tiempo y los sistemas distribuidos no pueden eliminar ese retardo

Indecibilidad de la partición

Es prácticamente imposible determinar con certeza si un nodo está completamente inaccesible debido a una falla de partición de la red o simplemente está tardando mucho en responder. Esta incertidumbre crea un dilema entre esperar por la consistencia o proceder para mantener la disponibilidad.


Compromiso de diseño

La arquitectura de un sistema distribuido debe hacer compromisos en función de su caso de uso. Por ejemplo, una base de datos para transacciones bancarias priorizará la consistencia, mientras que una aplicación de streaming de video priorizará la disponibilidad.



finalmente....

Entender el Teorema CAP no es solo un ejercicio académico; tiene aplicaciones prácticas y profundas en la ingeniería de sistemas actuales



Es importante porque...

Toma de decisiones informadas


Conocer las compensaciones entre Consistencia, Disponibilidad y Tolerancia a la Partición permite a los ingenieros tomar decisiones más informadas sobre la arquitectura y las tecnologías que se deben utilizar.

Adaptabilidad y escalabilidad

El teorema CAP da un marco para entender cómo un sistema se adaptará cuando se escale. ¿Podrá mantener su consistencia a medida que crece? ¿O tendrá que sacrificarla para mantener la disponibilidad?

Gestión del riesgo

El entendimiento de las implicaciones de CAP puede ayudar en la planificación de escenarios de falla y en la construcción de sistemas más resilientes.



Es importante porque...

Casos de uso Específico

No existe un 'mejor' sistema distribuido; todo depende del caso de uso. Algunos sistemas favorecen la consistencia sobre la disponibilidad, o viceversa, según las necesidades del negocio.

Ética y responsabilidad

La elección entre consistencia y disponibilidad puede tener implicaciones éticas. Por ejemplo, en aplicaciones de atención médica, la falta de consistencia en los datos podría ser fatal.

Resiliencia vs Eficiencia

A medida que avanzamos hacia un mundo cada vez más digital, el debate entre la resiliencia del sistema y la eficiencia en el rendimiento se vuelve más crítico, y el CAP proporciona un marco para ese debate



Esta diapositiva tiene como objetivo ofrecer una perspectiva completa y enriquecedora que resuma por qué el Teorema CAP es un tema tan relevante y cómo afecta múltiples aspectos de los sistemas distribuidos y sus aplicaciones.

