



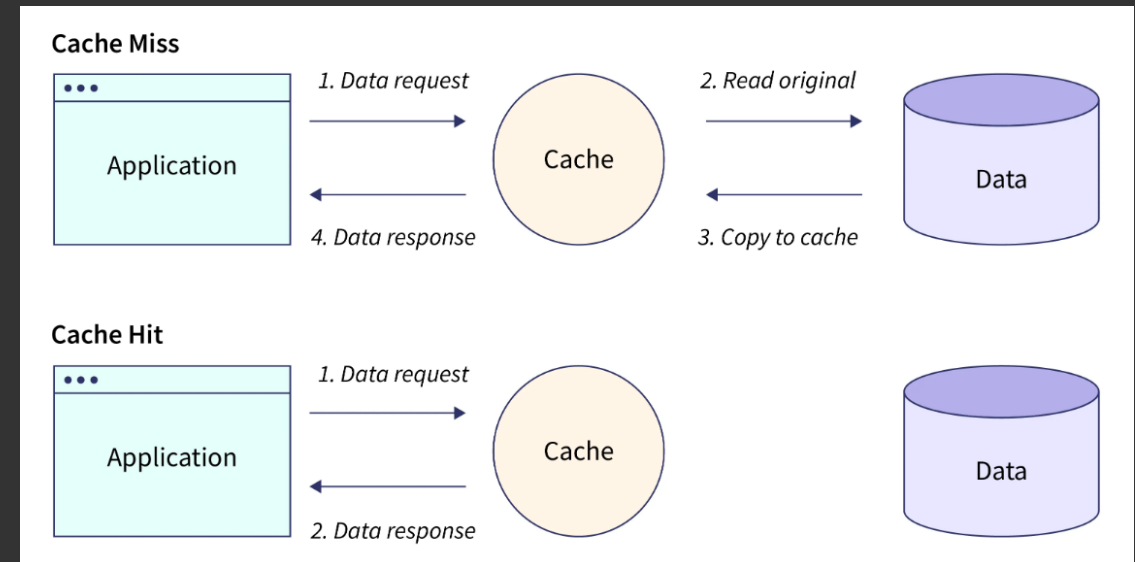
Caching

Introduction to Caching

What is Caching

Caching allows you to efficiently reuse previously retrieved or computed data.

In computing, a cache is a high-speed data storage layer which stores a subset of data, typically transient in nature, so that future requests for that data are served up faster than is possible by accessing the data's primary storage location.



Caching In Multiple Layers

- **DNS**: Domain to IP Resolution

When a user requests a website, their system or browser queries DNS servers to translate the domain name into an IP address. Caching these results means that repeated requests for the same domain are resolved faster, reducing the need for repetitive network lookups and improving overall browsing efficiency.

This is particularly beneficial for reducing latency and enhancing user experience by avoiding delays in domain resolution.

Caching In Multiple Layers

- **Web:** Accelerate retrieval of web content from web/app servers.

Web caching enhances performance by storing copies of frequently accessed web pages, images, and other content. When users request content, the cache delivers it quickly without needing to retrieve it from the origin server, which reduces load times and server strain. This not only speeds up content delivery but also reduces bandwidth usage and server costs. This is typically implemented using **CDN**.

Web caching is crucial for improving the responsiveness of web applications and ensuring a smoother, faster user experience.

Caching In Multiple Layers

- **Application:** Accelerate application performance and data access

Application caching improves performance by storing frequently accessed data and computational results closer to the application. This reduces the need for repeated and time-consuming database queries or complex calculations. This is usually implemented by **Memcached** or **Redis**.

This Caching is especially valuable for high-traffic applications, where it can significantly enhance scalability and reduce backend load.

Benefits of Caching

- It make data access faster and less expensive
- It improves the performance of the application.
- It gives responses quickly.
- Data access from memory is always faster than fetching from database.
- It reduces the costly backend requests.

Caching Terminologies

Understanding caching involves several key concepts that are crucial for effective implementation and management:

1. Cache Hit and Cache Miss
2. Cache Eviction Policies
LRU, LFU, FIFO, TTL based
3. Cache Expiry and TTL
4. Cache Loading Strategies
Lazy Loading, Eager Loading or Cache Warming
5. Cache size and capacity

