

What is the Caching Mechanism ?

Last Updated : 02 Jun, 2022



In this article, we will know about the Caching Mechanism, its importance, and basic usage, along with an understanding of its working mechanism, various types of caching & real-time applications.

What Is Caching?

It is the process of storing and accessing data from memory(i.e. cache memory). The main feature of caching is to reduce the time to access specific data. Caching aims at storing data that can be helpful in the future. The reason for caching is that accessing the data from persistent memory(hard drives like HDD, SSD) used to take considerable time, thus, slowing the process. Hence, caching reduces the time to acquire the data from memory. Cache memory is used to store the data which is a high-speed data storage layer and works for the purpose to reduce the need to access the data storage layer. Cache memory is implemented by fast access hardware (RAM).

Cache makes it possible to implement the mechanism for reusing previously computed data. Whenever hardware or software requests the specific data, the requested data is firstly searched in cache memory, if the data is found cache hit occurs and if data is not found cache miss occurs.

Why caching is important?

- Plays a vital role in improving system performance.
- Reduces the overall time and makes the system time efficient.
- Caching is unavoidable as it provides high performance in computer technology.
- It often does not make any new requests.
- It avoids reprocessing of the data.

Use Case of Caching Memory:

- Mainly caching is used to speed up database applications. Here, a portion of the database is replaced by cache and the result is removing latency which is used to come from frequently accessing the data. These use cases are seen in case of a larger volume of accessing the data, for instance, high traffic dynamic websites.
- Another use case can be an acceleration of query. Here, the cache is used for storing the result of a complex query. Queries such as the order and grouping take a considerable time for execution. If executing the queries is repeated then using cache for storing the result gives a higher response.

How caching mechanism works?

A partition is reserved in RAM for cache memory. Whenever a software requests data from

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our [Cookie Policy](#) & [Privacy Policy](#).

Got It !

exist in the cache memory, then, the application(software)will request the source(Hard Drives). After reading the data, the same data is stored in cache memory for future references.

As the cache memory is of limited size, the already existing data in the cache is to be removed for storing the new data. Caching system then requires an algorithm for removing the useless data. The algorithm will be designed to remove the data which is not going to get looked forward to soon.

For instance, [LRU\(least recently used\)](#) algorithm will remove the record which is least used

or demanded by the applications, acting on the principle, if the data is not used yet or might be not get used in the future. The challenge that comes with cache memory is "Cache misses", i.e. the request sent by the application for the data but results in absence of the data in cache memory. Too many Cache miss results in a decrease in efficiency. As the application's request for the data will be first checked in cache memory and if it results in a miss then the request is sent to the storage database (Hard drive) and extra work is done by the system, hence decreasing the performance. This challenge can be overcome by using large cache memory. Also Distributed Caches are commonly used for accessing the data which is bigger in amount. A distributed cache is the collective memory of RAM of more than one computer connected.

Types of Caching:

- **Database Caching:** A certain level of caching is already present in the Database. This internal cache is utilized for avoiding repeated queries. The last executed query result can be provided by the database immediately. The most commonly used database caching algorithm is based on storing key-value pairs in the hash table.
- **Memory-caching:** RAM is directly used for storing the cached data, this approach is faster than common database storage systems (Hard drives). This method is based on a set of key-value pairs in the database. The value is the cached data and the key is the unique value. Each set is uniquely identified. This Approach is Fast, efficient, and easy to implement.
- **Web caching:** It is divided into 2 parts:
 - **Web client caching:** This caching method is on the client-side and is commonly used by all internet users. Also known as web browser caching. It is started when the web page is loaded by the browser, it accumulates page resources like images, texts, media files, and scripts. When the same page is hit, the browser can grab the resources from the cache. It is faster than downloading from the internet.
 - **Web server caching:** Here, the resources are saved on the server-side aiming for the reuse of the resources. This approach is helpful in the case of dynamic web pages and this might not be useful in the case of static web pages. It reduces the overloading of the server, reduces the work, and increases the speed of page delivery.
- **CDN Caching:** Content delivery network caching is aimed at caching resources in proxy servers such as scripts, stylesheets, media files, and web pages. It acts as the gateway between the origin server and the user. When the user requests a resource, the proxy server comes into action to identify if it has a copy. If a copy of the found resource is carried to the user: else the request is processed by the origin server. It removes network latency and decreases the calling of the origin server.

Challenges with Caching:

- **Local Cache:** A cache coherence problem may occur as the locally cached resources are unevenly distributed from system to system, this process may slow down caching efficiency.
- **Cache Coherence problem:** When one of the processors modifies its local copy of data that is shared among several caches, keeping numerous local caches in synchronized order becomes a challenge.
- **Cache penetration:** When a user queries for data, the request is first sent to cache memory if the required data is present then it is sent to the user but if cache memory has no such data the request will be retransmitted to the main memory, and if the data is not present in the main memory it will result in NULL which causes cache penetration. This challenge can be overridden by the use of a bloom filter.
- **Cache avalanche:** If cache memory fails at the same time that a user requests a large query, the database will be put under a tremendous amount of stress, perhaps causing the database to crash. This challenge can be overcome by using cache cluster and

hystrix.

Advantages of Cache:

- Cache reduces the time taken to process a query, Cache minimizes the trip for requesting the same data.
- Cache decreases the load on the server.
- Cache increases the efficiency of system hardware.
- Web page downloading/rendering speed is increased with the help of caching.

Disadvantages of Cache:

- Cache algorithms are hard and complex to implement.
- Cache increases the complexity of an application.
- High maintenance cost.

Applications of Caching:

The caching mechanism is used in the following industries:

- **Health and wellness:** You'll be able to deliver quick speed, save overall spending, and expand as your usage grows with an effective caching architecture.
- **Advertising technology:** A millisecond can mean the difference between placing a bid on time and it becoming irrelevant when developing real-time bidding software. This necessitates a lightning-fast retrieval of bidding information from the database. Database caching, which can retrieve bidding data in milliseconds or less, is a terrific way to achieve that high level of speed.
- **Mobile:** Your mobile apps can achieve the performance your consumers need, grow exponentially, and decrease your total cost with appropriate caching solutions.
- **Gaming and media:** Caching helps to keep the game operating smoothly by delivering sub-millisecond query responses for commonly asked data.
- **Ecommerce:** Well-executed caching management is a strategic part of response time that can be the key differentiator, between making a sale and losing a customer.

 Comment

 More info

 Advertise with us

[Next Article >](#)

[What is DNS Caching](#)

Similar Reads

What is Pre-Caching?

Pre-caching is like getting ready for something before it happens. Imagine you're going on a trip and you pack your bag the night before so you're all set to go in the morning. That's pre-caching! In the digital world, it's when your device stores...

 13 min read

What is DNS Caching

DNS caching is a temporary storage system that keeps records of recent domain name lookups like google.com - 172.217.0.46 to speed up future requests. Instead of querying a DNS server every time you visit a website, your computer o...

 7 min read

What is Caching Strategies in DBMS?

In today's digital world, the speed of an application plays a major role in its success. Generally, users expect the applications to run faster with quick responses. Also, It should support seamless experiences across all their digital interactions, whether...

 9 min read

What is Result Caching in Oracle?

Result Caching is a feature of the Oracle database that allows query results to be stored in memory, eliminating the need to re-execute the query and improving overall query performance. It can be used in both single-instance and clustered...

 4 min read

What is Memcached?

Memcached is a powerful tool used in system design to speed up web applications by storing data in memory. It works as a caching layer, reducing the time needed to access frequently requested information. This helps websites and services handl...

🕒 12 min read

What is Cache Warming?

Cache warming is a proactive technique used in computing to pre-load data into a cache before it is actually needed by the system. This process helps to enhance the performance and speed of applications by reducing latency and ensuring that dat...

🕒 8 min read

📍 **Corporate & Communications Address:**
A-143, 7th Floor, Sovereign Corporate
Tower, Sector- 136, Noida, Uttar Pradesh
(201305)

📍 **Registered Address:**
K 061, Tower K, Gulshan Vivante
Apartment, Sector 137, Noida, Gautam
Buddh Nagar, Uttar Pradesh, 201305



Advertise with us

Company

Languages

DSA

Data Science &

Web

Python Tutorial

About Us
Legal
Privacy Policy
In Media
Contact Us
Advertise with us
GFG Corporate
Solution
Placement Training
Program

Python
Java
C++
PHP
GoLang
SQL
R Language
Android Tutorial
Tutorials Archive

Data Structures
Algorithms
DSA for Beginners
Basic DSA Problems
DSA Roadmap
Top 100 DSA
Interview Problems
DSA Roadmap by
Sandeep Jain
All Cheat Sheets

ML
Data Science With
Python
Data Science For
Beginner
Machine Learning
ML Maths
Data Visualisation
Pandas
NumPy
NLP
Deep Learning

Technologies
HTML
CSS
JavaScript
TypeScript
ReactJS
NextJS
Bootstrap
Web Design

Python Programming
Examples
Python Projects
Python Tkinter
Python Web Scraping
OpenCV Tutorial
Python Interview
Question
Django

Computer Science

Operating Systems
Computer Network
Database
Management System
Software Engineering
Digital Logic Design
Engineering Maths
Software
Development
Software Testing

DevOps

Git
Linux
AWS
Docker
Kubernetes
Azure
GCP
DevOps Roadmap

System Design

High Level Design
Low Level Design
UML Diagrams
Interview Guide
Design Patterns
OOAD
System Design
Bootcamp
Interview Questions

Interview Preparation

Competitive
Programming
Top DS or Algo for CP
Company-Wise
Recruitment Process
Company-Wise
Preparation
Aptitude Preparation
Puzzles

School Subjects

Mathematics
Physics
Chemistry
Biology
Social Science
English Grammar
Commerce
World GK

GeeksforGeeks

Videos

DSA
Python
Java
C++
Web Development
Data Science
CS Subjects