Caching

Course: Java

S1

© 2021 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.

Caching

Our application contains data for movies: cast, crew, box office numbers, reviews, etc.

We are adding in a new feature that allows people to find movies by searching two or more of the people starring in the film.

As an example, the user would input Johnny Depp and Orlando Bloom and our app would provide the result: Pirates of the Caribbean.

The Problem

This new feature will be great for our users. But how else can it be used in our app? For example: Can we use it to find out information for new movies, streaming releases, or movie stars whose names are in the news?

What can we do to make sure that the processing needs of this new feature don’t reduce the load time for common requests?

Learning Outcomes

By the end of this lesson, you will be able to:

- Explain what a cache is.
- Compare and contrast buffering and caching.
- Discuss when to use a cache.
- Use a cache in a microservices architecture.

What is a Cache?

Cache

What is a cache?

01

Stores data for future reference

02

A copy of the original data

03

Improves retrieval time

A website caches images to improve performance, specifically images that may appear on multiple pages of the site.

Why Use a Cache?

Why Use a Cache?

It’s less expensive:

- Saves money
- Saves processing time

Why Use a Cache?

Data transformation requires processing power. Once that data is transformed, caching the results saves our app from having to process it again.

Why Use a Cache?

Improved user experience.

Compare and Contrast Buffering and Caching

When a video app is loading an .mp4 file, the file is buffered to match the transmission speed of the sender and the receiver.

Like caching, buffering is useful for improving the user experience.

Buffering preloads data from the “original” source.

Caching is accessing a copy of the data.

How Does a Cache Work?

01 A user opens an app.

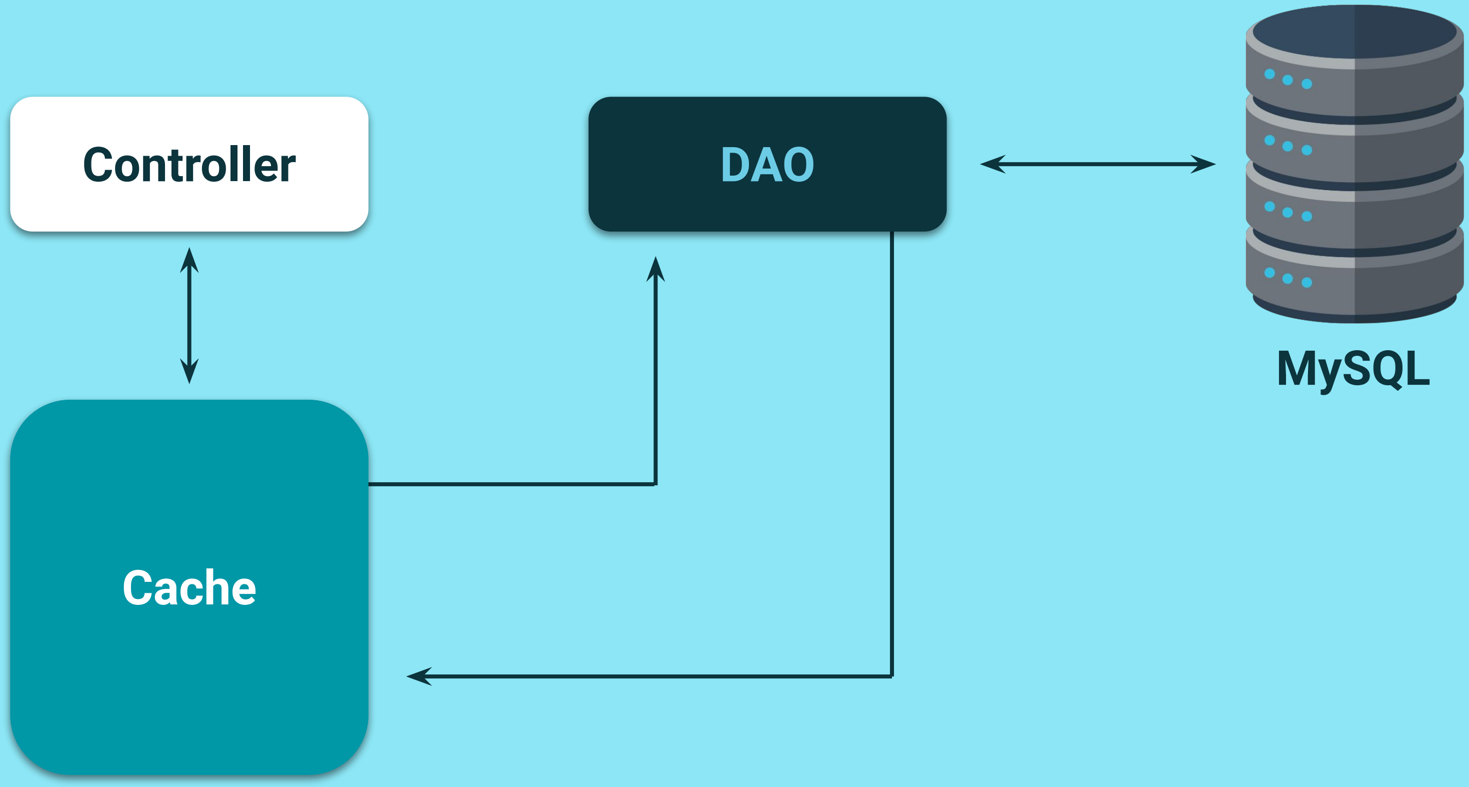
02 The app attempts to fetch data from the cache.

03 If the data exists in the cache, it is presented.

04 If the data is not in the cache, the data is requested from the database.

05 This data may then be cached for faster retrieval when future requests are made.

How Does a Cache Work? (Diagram)



Things to Consider When Using a Cache

Caches take up disk space.

Cached data may not be the most recent, therefore not accurate.

Caches are most useful with generic data, ie. not user specific.

When data changes often, a cache is likely not very effective.

If the accuracy of the data is critical, a cache is likely not the best solution.

Activity: Problem Discussion

In this activity you will work with your group and discuss and whiteboard potential solutions to the problem.

Suggested Time:

5 minutes

RECAP

Learning Outcomes

By the end of this lesson, you will be able to:

- Explain what a cache is.
- Compare and contrast buffering and caching.
- Discuss when to use a cache.
- Use a cache in a microservices architecture.

Questions?