

Homework 8

due October 31, 2024

Caching

One way in which an application can interact with a cache is the **cache-aside pattern**. We illustrate this pattern in Figure 1. The application controls the behavior of the cache. If the application writes data, it updates both the database and the cache. If the application reads data, it first tries the cache. If the item is not found in the cache, the application consults the database and also updates the cache. Items in the cache are either removed or invalidated if their Time-To-Live (TTL) is larger than a given threshold. In this case, the next access to the item results in a cache-miss and the application will update the cache with a newer copy of the item.

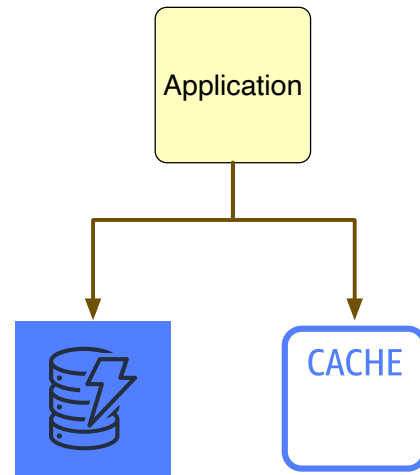


Figure 1: Cache-aside pattern

Other patterns have the application only interact with the cache. In this case, we distinguish between the following patterns.

Read-through:

The application satisfies all read requests by accessing the cache. If the data required is not available in the cache, a loader is invoked to access the data systems and load the results in the cache for the application to utilize. This pattern is always used.

Write-through:

The application always writes updates to the cache. When the application updates the cache the cache automatically writes the new cache values to the database. When the database is updated, the application can complete the request.

Write-behind / Write back:

Like write-through, except the application does not wait for the value to be written to the database from the cache. This increases request responsiveness at the expense of possible lost updates if the cache server crashes before a database update is completed.

Questions:

- (1) Read the MS Azure documentation on the cache-aside pattern. List the situation in which the cache-aside pattern is justified.

- (2) What is the (client-centric) consistency model when using the cache-aside pattern? Here we assume that an application only interacts with one replica of the cache, but that other applications can have their own cache.
- (3) What are the client-centric consistency models that apply to a write-through cache that is shared among all applications?

HTTP Headers

Look up and shortly explain the Cache Control and Expired and Modified HTTP headers. How does etag work?