Sistemas de Computação em Cloud

2024/2025

Project 1
TuKano

Authors:

70525, Francisco Silva
70596, Martim Latas

Practical class Nº 3

November 9th, 2024

Introduction to the Project

In this project, we were tasked with the provided code for the web application *TuKano*, adapting it to run on the Microsoft Azure Cloud platform. *TuKano* is a social network for sharing short videos, inspired by popular platforms like TikTok and YouTube Shorts. Our task involves transforming this single-server application using the file system into a cloud solution by integrating Azure's Platform as a Service (PaaS).

The modified *TuKano* architecture we implemented incorporates Azure Cosmos DB for storing the blobs from the shorts, and Azure Cache for Redis to enhance data retrieval speed.

When deciding on how to implement this Cache, we decided to store the most recent data on a list. Midway through our code, we found out that it would be best if we actually didn't store the likes and follows on cache, since they're mostly relational and when retrieving the likes, for example, of a short, it is probable that not every like is stored in the cache. This would lead into eventual calls to the database, whether it would be for a portion of the data or all of it, so we went against storing Likes and Follows on cache.

When adding a new user or short, it will also be inserted in the cache, as expected. This would make it so that an eventual GET operation would result in a faster runtime, effectively making our application run smoother when there would be a lot of operations coming in at the same time.

When performing GET operations on Users or Shorts, our application first checks if the intended object is on cache (if cache is on at the moment). On a successful check, it will return the value obtained from cache, and if it fails, then it has to get the data from the database.

When updating these (existing) objects, and the cache is turned on, then the application has to update the value it has stored, else there would be some conflicting data coming from GETs and POSTs. This operation can be called either way, since if the intended value is not on cache then nothing would happen.

We connected our blob file system to the Azure Cosmos DB via the CloudsystemStorage.java class. Here, we connect to the shorts container in our Azure platform and give the appropriate connection string. In this class we set up the methods to write and read from the BlobContainerClient object that reads and writes to the appropriate "directories," given a path to a blob.

On the rest of the java classes, there was not much room for improvement as we still used the Hibernate + Postgre given on the source code, although we did

have to change the way some queries were made, as some FROMs were necessary to delete the query results.

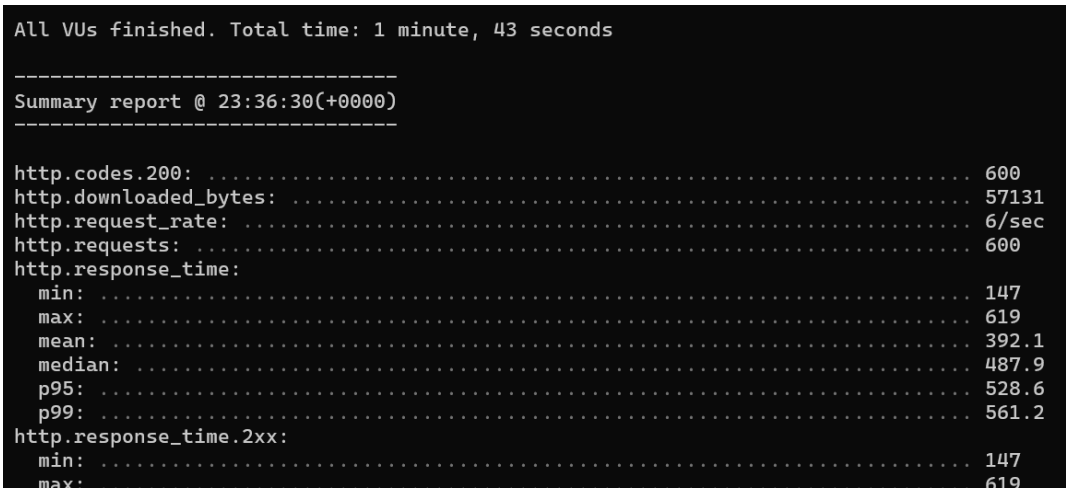
To use PostgreSQL in this setup, we created an Azure Cosmos DB for PostgreSQL cluster, which enables persistent data storage. This setup allows data to remain available even if the Docker container used for the local hosting is reset or restarted, ensuring that no data is lost during these events, and that we can use the SQL queries that have been given us from the base project code.

Performance Analysis via Artillery

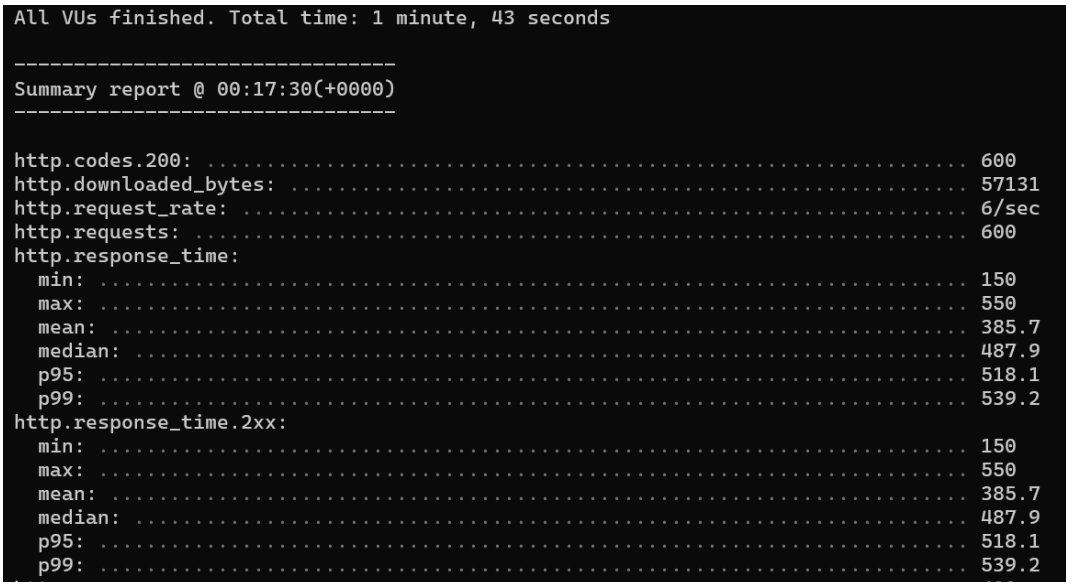
Although we did implement our own tests on artillery, we decided to test the ones given to us by the teachers, since ours were simpler, made more to test specific scenarios/operations (as requested), while the ones given to us simulate more accurately the way this application would behave in the real world.

Register Users

With Cache:



Without Cache:



Using caching typically results in improved response time metrics. The highest response time was 619 ms, while the average value was 392.1 ms. In the absence of caching, the overall response time metrics are somewhat decreased. The highest response time recorded was 550 ms, while the mean stood at 385.7 ms.

The information suggests that caching does not offer a noteworthy performance improvement in this situation. In reality, with caching, response times

can be marginally longer in the worst scenarios, raising the average. This rise most likely results from the additional process of saving the newly created user in both the database and the cache.

Upload Shorts

With Cache:

```
All VUs finished. Total time: 11 seconds

--------------------------------
Summary report @ 23:48:58(+0000)
--------------------------------

http.codes.200: ..................................................................... 30
http.codes.204: ..................................................................... 30
http.downloaded_bytes: .............................................................. 8675
http.request_rate: .................................................................. 6/sec
http.requests: ...................................................................... 60
http.response_time:
  min: .............................................................................. 125
  max: .............................................................................. 710
  mean: ............................................................................. 405.1
  median: ........................................................................... 415.8
  p95: .............................................................................. 685.5
  p99: .............................................................................. 699.4
http.response_time.2xx:
  min: .............................................................................. 125
  max: .............................................................................. 710
  mean: ............................................................................. 405.1
  median: ........................................................................... 415.8
  p95: .............................................................................. 685.5
  p99: .............................................................................. 699.4
```

Without Cache:

```
All VUs finished. Total time: 11 seconds

--------------------------------
Summary report @ 00:21:36(+0000)
--------------------------------

http.codes.200: ..................................................................... 30
http.codes.204: ..................................................................... 30
http.downloaded_bytes: .............................................................. 8632
http.request_rate: .................................................................. 5/sec
http.requests: ...................................................................... 60
http.response_time:
  min: .............................................................................. 124
  max: .............................................................................. 689
  mean: ............................................................................. 410.8
  median: ........................................................................... 487.9
  p95: .............................................................................. 685.5
  p99: .............................................................................. 685.5
http.response_time.2xx:
  min: .............................................................................. 124
  max: .............................................................................. 689
  mean: ............................................................................. 410.8
  median: ........................................................................... 487.9
  p95: .............................................................................. 685.5
  p99: .............................................................................. 685.5
```

The response time metrics reveal notable performance improvements due to caching, particularly in the average response time, which decreased from 410.8 ms without caching to 405.1 ms with caching. While minimum response times remain nearly identical, indicating minimal impact on optimal latency, average performance is clearly enhanced. The maximum response time is slightly higher with caching (710 ms) but remains comparable to the non-cached version.

Caching also positively affects the HTTP request rate, allowing the system to handle more requests per second by reducing response times. This improved request rate suggests that caching enables the system to process requests more efficiently within the same timeframe.

Realistic Flow

With Cache:

First Flow -

```
All VUs finished. Total time: 11 seconds

--------------------------------
Summary report @ 23:50:43(+0000)
--------------------------------

http.codes.200: ..................................................................... 16
http.codes.204: ..................................................................... 8
http.codes.500: ..................................................................... 10
http.downloaded_bytes: .............................................................. 367401
http.request_rate: .................................................................. 4/sec
http.requests: ...................................................................... 34
http.response_time:
  min: .............................................................................. 177
  max: .............................................................................. 871
  mean: ............................................................................. 338.7
  median: ........................................................................... 314.2
  p95: .............................................................................. 528.6
  p99: .............................................................................. 550.1
http.response_time.2xx:
  min: .............................................................................. 177
  max: .............................................................................. 871
  mean: ............................................................................. 363.4
  median: ........................................................................... 327.1
  p95: .............................................................................. 528.6
  p99: .............................................................................. 550.1
http.response_time.5xx:
  min: .............................................................................. 260
  max: .............................................................................. 341
  mean: ............................................................................. 279.6
  median: ........................................................................... 267.8
  p95: .............................................................................. 290.1
  p99: .............................................................................. 290.1
```

Second Flow -

```
All VUs finished. Total time: 11 seconds

--------------------------------
Summary report @ 00:05:03(+0000)
--------------------------------

http.codes.200: ..................................................................... 17
http.codes.204: ..................................................................... 9
http.codes.500: ..................................................................... 11
http.downloaded_bytes: .............................................................. 636739
http.request_rate: .................................................................. 5/sec
http.requests: ...................................................................... 37
http.response_time:
  min: .............................................................................. 112
  max: .............................................................................. 586
  mean: ............................................................................. 323.9
  median: ........................................................................... 278.7
  p95: .............................................................................. 550.1
  p99: .............................................................................. 561.2
http.response_time.2xx:
  min: .............................................................................. 112
  max: .............................................................................. 586
  mean: ............................................................................. 345.2
  median: ........................................................................... 295.9
  p95: .............................................................................. 550.1
  p99: .............................................................................. 561.2
http.response_time.5xx:
  min: .............................................................................. 263
  max: .............................................................................. 294
  mean: ............................................................................. 273.6
  median: ........................................................................... 273.2
  p95: .............................................................................. 278.7
  p99: .............................................................................. 278.7
```

Without Cache:

First Flow -

```
All VUs finished. Total time: 11 seconds

--------------------------------
Summary report @ 00:23:16(+0000)
--------------------------------

http.codes.200: ..................................................................... 25
http.codes.204: ..................................................................... 5
http.codes.500: ..................................................................... 8
http.downloaded_bytes: .............................................................. 592980
http.request_rate: .................................................................. 5/sec
http.requests: ...................................................................... 38
http.response_time:
  min: .............................................................................. 139
  max: .............................................................................. 632
  mean: ............................................................................. 332.4
  median: ........................................................................... 278.7
  p95: .............................................................................. 596
  p99: .............................................................................. 608
http.response_time.2xx:
  min: .............................................................................. 139
  max: .............................................................................. 632
  mean: ............................................................................. 345.4
  median: ........................................................................... 327.1
  p95: .............................................................................. 596
  p99: .............................................................................. 608
http.response_time.5xx:
  min: .............................................................................. 263
  max: .............................................................................. 341
  mean: ............................................................................. 283.4
  median: ........................................................................... 273.2
  p95: .............................................................................. 290.1
  p99: .............................................................................. 290.1
```

Second Flow -

```
All VUs finished. Total time: 11 seconds

--------------------------------
Summary report @ 00:26:19(+0000)
--------------------------------

http.codes.200: ..................................................................... 22
http.codes.204: ..................................................................... 6
http.codes.500: ..................................................................... 8
http.downloaded_bytes: .............................................................. 505704
http.request_rate: .................................................................. 4/sec
http.requests: ...................................................................... 36
http.response_time:
  min: .............................................................................. 133
  max: .............................................................................. 625
  mean: ............................................................................. 329.3
  median: ........................................................................... 278.7
  p95: .............................................................................. 561.2
  p99: .............................................................................. 584.2
http.response_time.2xx:
  min: .............................................................................. 133
  max: .............................................................................. 625
  mean: ............................................................................. 341.3
  median: ........................................................................... 278.7
  p95: .............................................................................. 561.2
  p99: .............................................................................. 584.2
http.response_time.5xx:
  min: .............................................................................. 274
  max: .............................................................................. 303
  mean: ............................................................................. 287.1
  median: ........................................................................... 278.7
  p95: .............................................................................. 295.9
```

In the second test utilizing cache, the request rate achieved 5 requests per second, surpassing both the first test with cache and the one without, suggesting that the cache may assist in managing a greater load, enhancing the system's ability to process more requests within the same timeframe.

In this instance, the average response time remains fairly uniform throughout all tests. Nonetheless, the minimum response time demonstrates a noteworthy enhancement with cache, particularly in the second test (112 ms), in contrast to the non-cache situation (133 ms). This indicates that cache decreases the minimum latency, probably by handling repeated requests faster.

The highest response time is greater without cache (625 ms) in comparison to with cache. The increase in the highest response time during the initial test with cache may suggest an anomaly or outlier that requires additional examination.

The steady mean response time for responses with a 200 status code throughout the tests indicates reliable performance. Additionally, the slightly improved 95th and 99th percentile response times with caching suggest enhanced handling of outliers and peak demand scenarios.

Delete Users

With Cache:

```
All VUs finished. Total time: 1 minute, 41 seconds

--------------------------------
Summary report @ 00:10:34(+0000)
--------------------------------

http.codes.200: ..................................................................... 200
http.downloaded_bytes: .............................................................. 27367
http.request_rate: .................................................................. 2/sec
http.requests: ...................................................................... 200
http.response_time:
  min: .............................................................................. 472
  max: .............................................................................. 719
  mean: ............................................................................. 503.5
  median: ........................................................................... 497.8
  p95: .............................................................................. 528.6
  p99: .............................................................................. 584.2
http.response_time.2xx:
  min: .............................................................................. 472
  max: .............................................................................. 719
  mean: ............................................................................. 503.5
  median: ........................................................................... 497.8
  p95: .............................................................................. 528.6
  p99: .............................................................................. 584.2
```

Without Cache:

```
All VUs finished. Total time: 1 minute, 41 seconds

--------------------------------
Summary report @ 00:32:01(+0000)
--------------------------------

http.codes.200: ..................................................................... 200
http.downloaded_bytes: .............................................................. 27367
http.request_rate: .................................................................. 2/sec
http.requests: ...................................................................... 200
http.response_time:
  min: .............................................................................. 468
  max: .............................................................................. 640
  mean: ............................................................................. 501.7
  median: ........................................................................... 497.8
  p95: .............................................................................. 528.6
  p99: .............................................................................. 596
http.response_time.2xx:
  min: .............................................................................. 468
  max: .............................................................................. 640
  mean: ............................................................................. 501.7
  median: ........................................................................... 497.8
  p95: .............................................................................. 528.6
  p99: .............................................................................. 596
```

The average response time for delete user functions with caching turned on is 503.5 ms, whereas the average without caching is a bit lower at 501.7 ms. This slight variation indicates that activating cache has a minimal effect on the average

response time for delete operations, which are probably less reliant on cache by nature.

Conversely, the highest response time recorded with caching (719 ms) is significantly greater than that without caching (640 ms). This discrepancy might stem from cache-related overhead, including validation or synchronization lags, particularly during peak operating periods. This observation suggests that caching may cause variation in latency, potentially affecting the highest response times.

The information shows that caching does not offer a notable performance increase in this situation. In reality, with caching, the response times are a bit longer in the worst scenarios, raising the average. This rise probably results from the additional step of checking whether the user is in the cache, and if so, also requiring deletion to ensure consistency.

Conclusion

We have come to the conclusion that the TuKano project is an app where content is often accessed and read-heavy, which means that it should be implemented with the use of caching.

With the results of the tests above, even if adding a new resource to the database does take some extra time, the most common operations would be requesting values from the database, whether they would be users or shorts.

If the most common values would be gets, then the cache would, obviously, help greatly with the response times of the system, as some (or even most) calls, depending on the storage of this cache, don’t need to go to the “main” application and can be retrieved sooner, without causing that many errors on the way due to conflicts in the requests.