

Cloud Computing Project 1

Kacper Kuźnik 75267

Mikołaj Nowacki 75231

Dawid Bogacz 75160

Introduction

This project is a cloud-based web application for managing LEGO collections and online auctions, developed in Python using FastAPI and Microsoft Azure services. It provides a backend system that allows users to register accounts, upload LEGO sets with photos, comment on sets, and participate in auctions. Data is stored in Azure Cosmos DB, while images are handled through Azure Blob Storage. Redis caching is used to enhance performance and reduce query latency. Additionally, the system includes automated background tasks implemented with Azure Functions, such as closing auctions and analyzing comment sentiment using TextBlob. Based on these analyses, the platform periodically computes the most liked LEGO sets. The project demonstrates the design of a scalable, event-driven cloud architecture combining data management, serverless computing, and natural language processing, providing a solid foundation for a real-world collectible marketplace application.

Design

The system is designed as a modular, cloud-based architecture built around FastAPI as the core backend framework. It follows a service-oriented approach, integrating multiple Microsoft Azure components to ensure scalability, reliability, and performance. The main components include Azure Cosmos DB for data storage, Azure Blob Storage for managing images, Azure Functions for background tasks, and Redis for caching frequently accessed data. User accounts, LEGO sets, comments, and auctions are each represented as distinct containers within Cosmos DB, enabling structured yet flexible data handling. When users upload LEGO sets, the images are stored in Blob Storage and linked to the database through secure URLs. The interaction between services is event-driven. In example, Azure Functions automatically process scheduled computations, such as determining which auctions should be closed. This decoupled, asynchronous design ensures smooth performance and simplifies the deployment of new features without affecting existing functionality.

Implementation

The backend implementation uses FastAPI to define a RESTful API exposing endpoints for managing users, LEGO sets, comments, and auctions. Each endpoint interacts directly with Cosmos DB containers via the Azure Cosmos Python SDK, using partition keys for efficient queries. User passwords are securely hashed before storage, and all data operations are validated using Pydantic models. The image upload pipeline is managed through a custom

BlobStorageManager class, which handles file uploads and retrievals. Caching mechanisms use Redis to store query results, improving response times for frequently accessed endpoints. Serverless automation is achieved with Azure Functions, where scheduled triggers execute tasks like closing expired auctions or analyzing sentiment in user comments using TextBlob. The system is fully containerized, enabling easy deployment to Azure App Services. Overall, the implementation highlights efficient API design, clean modularization, and seamless integration of serverless and data-driven components.

Evaluation

a) Local

Surprisingly, local performance without caching is slightly better in terms of mean response time (389.3ms vs 506.6ms). Gives mixed results

b) Spain

Caching reduces mean and median response times by roughly 30–40% (3740ms vs 5399ms).

Maximum response times drop significantly with caching, but some endpoints like `/rest/legoset` remain slow (min 2964ms, max 7072ms).

`/rest/user` is faster with caching (mean 3789ms vs 7101ms without caching).

Caching is effective in reducing latency when the app is deployed in SpainCentral, but overall response times are still much higher than local, likely due to regional deployment latency.

c) Poland

Performance is worse than SpainCentral, despite caching.

High mean and median indicate network latency from the client to PolandCentral is dominating.

Endpoints `/rest/user` and `/rest/legoset/recent` have extremely high response times (up to 9.2s).

Deploying in PolandCentral introduces significant latency, due to geographical distance from test client and origin of the database. Caching cannot fully compensate for regional network delays.

Below are presented results of tests from artillery:

Local with caching

```
http.codes.200: ..... 186
http.downloaded_bytes: ..... 165283528
http.response_time:
  min: ..... 55
  max: ..... 5192
  mean: ..... 506.6
  median: ..... 165.7
  p95: ..... 1790.4
  p99: ..... 4676.2
http.response_time.2xx:
  min: ..... 55
  max: ..... 5192
  mean: ..... 506.6
  median: ..... 165.7
  p95: ..... 1790.4
  p99: ..... 4676.2
http.responses: ..... 186
plugins.metrics-by-endpoint./rest/legoset.codes.200: ..... 25
plugins.metrics-by-endpoint./rest/legoset/recent.codes.200: ..... 31
plugins.metrics-by-endpoint./rest/user.codes.200: ..... 130
plugins.metrics-by-endpoint.response_time./rest/legoset:
  min: ..... 714
  max: ..... 5192
  mean: ..... 2085
  median: ..... 1176.4
  p95: ..... 4676.2
  p99: ..... 5065.6
plugins.metrics-by-endpoint.response_time./rest/legoset/recent:
  min: ..... 55
  max: ..... 1228
  mean: ..... 347.8
  median: ..... 267.8
  p95: ..... 982.6
  p99: ..... 1085.9
plugins.metrics-by-endpoint.response_time./rest/user:
  min: ..... 105
  max: ..... 2779
  mean: ..... 240.9
  median: ..... 147
  p95: ..... 415.8
  p99: ..... 1525.7
vusers.completed: ..... 300
vusers.created: ..... 300
vusers.created_by_name.Create LegoSet: ..... 114
vusers.created_by_name.Create User: ..... 100
vusers.created_by_name.Get recent legosets: ..... 31
vusers.created_by_name.List LegoSets: ..... 25
vusers.created_by_name.List Users: ..... 30
vusers.failed: ..... 0
vusers.session_length:
  min: ..... 4.9
  max: ..... 5252.3
  mean: ..... 538.9
  median: ..... 347.3
  p95: ..... 1249.1
  p99: ..... 4492.8
```

Local without caching

```
http.codes.200: ..... 201
http.downloaded_bytes: ..... 187860515
http.response_time:
  min: ..... 108
  max: ..... 1705
  mean: ..... 389.3
  median: ..... 232.8
  p95: ..... 1043.3
  p99: ..... 1465.9
http.response_time.2xx:
  min: ..... 108
  max: ..... 1705
  mean: ..... 389.3
  median: ..... 232.8
  p95: ..... 1043.3
  p99: ..... 1465.9
http.responses: ..... 201
plugins.metrics-by-endpoint./rest/legoset.codes.200: ..... 30
plugins.metrics-by-endpoint./rest/legoset/recent.codes.200: ..... 30
plugins.metrics-by-endpoint./rest/user.codes.200: ..... 141
plugins.metrics-by-endpoint.response_time./rest/legoset:
  min: ..... 494
  max: ..... 1705
  mean: ..... 849.6
  median: ..... 685.5
  p95: ..... 1408.4
  p99: ..... 1465.9
plugins.metrics-by-endpoint.response_time./rest/legoset/recent:
  min: ..... 154
  max: ..... 822
  mean: ..... 273.9
  median: ..... 237.5
  p95: ..... 441.5
  p99: ..... 441.5
plugins.metrics-by-endpoint.response_time./rest/user:
  min: ..... 108
  max: ..... 1479
  mean: ..... 315.9
  median: ..... 198.4
  p95: ..... 804.5
  p99: ..... 1130.2
vusers.completed: ..... 300
vusers.created: ..... 300
vusers.created_by_name.Create LegoSet: ..... 99
vusers.created_by_name.Create User: ..... 104
vusers.created_by_name.Get recent legosets: ..... 30
vusers.created_by_name.List LegoSets: ..... 30
vusers.created_by_name.List Users: ..... 37
vusers.failed: ..... 0
vusers.session_length:
  min: ..... 3.8
  max: ..... 1778
  mean: ..... 464.1
  median: ..... 487.9
  p95: ..... 1085.9
  p99: ..... 1495.5
```

App deployed in region SpainCentral with caching

http.response_time:

min: 784
max: 7072
mean: 3740.8
median: 3197.8
p95: 6702.6
p99: 6702.6

http.response_time.2xx:

min: 784
max: 7072
mean: 3740.8
median: 3197.8
p95: 6702.6
p99: 6702.6

plugins.metrics-by-endpoint.response_time./rest/legoset:

min: 2964
max: 7072
mean: 5211.5
median: 4770.6
p95: 6064.7
p99: 6064.7

plugins.metrics-by-endpoint.response_time./rest/legoset/recent:

min: 784
max: 2510
mean: 1634.7
median: 1620
p95: 1620
p99: 1620

plugins.metrics-by-endpoint.response_time./rest/user:

min: 1541
max: 6691
mean: 3789.2
median: 3262.4
p95: 5598.4
p99: 5598.4

App deployed in SpainCentral without caching

http.response_time:

min: 247
max: 8924
mean: 5399.9
median: 6312.2
p95: 8692.8
p99: 8692.8

http.response_time.2xx:

min: 247
max: 8924
mean: 5399.9
median: 6312.2
p95: 8692.8
p99: 8692.8

http.responses: 21

plugins.metrics-by-endpoint.response_time./rest/legoset/recent:

min: 247
max: 6300
mean: 3131.8
median: 3828.5
p95: 4583.6
p99: 4583.6

plugins.metrics-by-endpoint.response_time./rest/user:

min: 1648
max: 8924
mean: 7101
median: 7865.6
p95: 8692.8
p99: 8692.8

App deployed in PolandCentral with caching

http.response_time:

min: 6727
max: 9283
mean: 8549.2
median: 9230.4
p95: 9230.4
p99: 9230.4

http.response_time.2xx:

min: 6727
max: 9283
mean: 8549.2
median: 9230.4
p95: 9230.4
p99: 9230.4

plugins.metrics-by-endpoint.response_time./rest/legoset/recent:

min: 9283
max: 9283
mean: 9283
median: 9230.4
p95: 9230.4
p99: 9230.4

plugins.metrics-by-endpoint.response_time./rest/user:

min: 6727
max: 9235
mean: 8365.8
median: 8352
p95: 9230.4
p99: 9230.4

Conclusions

The project successfully demonstrates the design and deployment of a scalable cloud-based web application using Microsoft Azure services. Through the integration of FastAPI, Cosmos DB, Blob Storage, Redis, and Azure Functions, the system effectively combines data management, serverless automation, and performance optimization within a modular architecture.

Performance testing revealed that caching significantly improves response times - particularly in remote deployments such as SpainCentral - reducing latency by up to 40%. However, results also show that geographical location plays a crucial role in overall performance, with local deployments offering the lowest latency. While Redis caching improves throughput, it cannot fully mitigate delays caused by network distance, as observed in the PolandCentral region.

To further enhance global performance and reduce latency for users distributed across regions, the next step should be the integration of a Content Delivery Network. A CDN would enable faster content delivery by caching static assets at edge locations closer to end users, minimizing network travel distance and improving user experience.