



Event-Driven cloud Architecture, CQRS and Event Sourcing for User Management

2019/08/09

About

Company	Molo17 S.r.l.
Author	Lorenzo Busin
State	Approved
Use	Internal
Email	lorenzo.busin@gmail.com

Description

Documentation about cloud architectures that use an Event-Driven approach and implement CQRS and Event Sourcing for User Management.

Table of contents

1	Introduction	4
1.1	Project purpose	4
1.2	References	4
2	Amazon Web Services	5
2.1	Lambda	5
2.1.1	Write mode Lambda functions	5
2.1.1.1	PushOperationAggregateToSQS	5
2.1.1.2	CommandOperationAggregate	6
2.1.1.3	Mediator	8
2.1.1.4	OperationAggregate	9
2.1.1.5	Recovery	10
2.1.2	Read mode Lambda functions	10
2.1.2.1	ReadOperationAggregate	10
2.1.3	CloudWatch Logs	11
2.2	DynamoDB	11
2.3	API Gateway	12
2.4	Simple Queue Service	12
3	Serverless Framework	13
3.1	Description	13
3.2	Serverless.yml	13
3.3	Project's root	16
4	CQRS	17
4.1	Architecture overview	18
4.2	Write model	19
4.3	Read model	19
5	User Management	20
5.1	Aggregates	20
5.1.1	User	20
5.1.2	Role	20
5.1.3	Authorization	20
5.1.4	Group	20
5.2	Admin side	21
5.2.1	Authentication	21
5.2.2	Use cases	21
5.3	User side	22
5.3.1	Authentication	22
5.3.2	Use cases	23
6	Event Sourcing	25
6.1	Event store	25
6.1.1	Event object	25

7	Extension points	27
7.1	New aggregates	27
7.2	New write functions	27
7.3	New read functions	27
8	Setup	28

List of figures

1	Fetch POST API	5
2	DynamoDB event object	8
3	Parsed DynamoDB event object	9
4	Fetch GET API	10
5	Plugins - serverless.yml	13
6	Provider info - serverless.yml	14
7	DynamoDB tables - serverless.yml	14
8	Lambda triggered by SQS event - serverless.yml	15
9	Lambda triggered by GET endpoint - serverless.yml	15
10	Project's root	16
11	Architecture overview	18
12	Write model sequence diagram	19
13	Read model sequence diagram	19
14	Admin use cases	21
15	User use cases	23
16	DynamoDB eventStore example item	26

1 Introduction

1.1 Project purpose

This project explains how to build Event-Driven architectures, CQRS and Event Sourcing for user management, showing how should be implemented, extended, strengths and weaknesses. The application for user management implements an authentication function and the CRUD operations for each aggregates, which are: users, roles, authorizations and groups.

1.2 References

- **What do you mean by "Event-Driven"? - Martin Fowler:**
martinfowler.com/articles/201701-event-driven.html;
- **CQRS - Martin Fowler:**
martinfowler.com/bliki/CQRS.html;
- **Event Sourcing - Martin Fowler:**
martinfowler.com/eaaDev/EventSourcing.html;
- **AWS Lambda documentation:**
<https://docs.aws.amazon.com/lambda/index.html>;
- **AWS DynamoDB documentation:**
<https://docs.aws.amazon.com/dynamodb/index.html>;
- **AWS API gateway documentation:**
<https://docs.aws.amazon.com/apigateway/index.html>;
- **AWS SQS documentation:**
<https://docs.aws.amazon.com/sqs/index.html>;
- **Serverless Framework:**
<https://serverless.com/>.

2 Amazon Web Services

Amazon Web Services (AWS) is a subsidiary of Amazon that provides on-demand cloud computing platforms to individuals, companies, and governments, on a pay-as-you-go basis. These cloud computing web services provide a set of primitive abstract technical infrastructure and distributed computing building blocks and tools. AWS's version of virtual computers emulate most of the attributes of a real computer including, hardware central processing units and graphics processing units, local/RAM memory, hard-disk/SSD storage; a choice of operating systems; networking; and pre-loaded application software such as web servers and databases.

2.1 Lambda

AWS Lambda is an event-driven, serverless computing platform provided by Amazon. It is a computing service that runs code in response to events and automatically manages the computing resources required by that code. The purpose of Lambda is to simplify building smaller, on-demand applications that are responsive to events and new information. AWS targets starting a Lambda instance within milliseconds of an event. Node.js, Python, Java, Go, Ruby and C# through .NET Core are all officially supported.

In this project, lambda functions are the computing core for the execution of all commands and are written in Node.js.

2.1.1 Write mode Lambda functions

2.1.1.1 PushOperationAggregateToSQS

Those functions starts the flow and triggered when fetching the corresponding API gateway URL with a POST request like this:

```
fetch("https://domain/resourceAPIpath", {
  method: "post",
  headers: {
    'Accept': 'application/json',
    'Content-Type': 'application/json'
  },
  body: JSON.stringify({
    "attribute1": "value1",
    "attribute2": "value2"
  })
});
```

Fig. 1: Fetch POST API

Once triggered, these functions retrieve the event and put it in the *MessageBody* parameter. The event object is not changed, instead in the user functions which encrypt the password before sending the message to corresponding SQS queue.

Example:

```

1 module.exports.pushUpdateRoleToSQS = (event, context, callback) => {
2   const AWS = require('aws-sdk');
3   const SQS = new AWS.SQS();
4   const stringedEvent = JSON.stringify(event);
5
6   const params = {
7     MessageBody: stringedEvent,
8     QueueUrl: "https://yourDomain/updateRoleQueue"
9   };
10
11  SQS.sendMessage(params, function(err, data){
12    if(err){
13      console.log(err);
14      callback(null, err);
15    }
16    else
17      callback(null, "Role event pushed to SQS");
18  });
19 };

```

2.1.1.2 CommandOperationAggregate

These functions validate the values of the attributes before storing the event in the *eventStore* and triggered when a new event arrives to the corresponding queue.

If you have to check for duplicated attributes in the database, the function must be marked *async* because it has to wait for the result of the check operation.

Example:

```

1 module.exports.commandCreateRole = async (event, context, callback) => {
2   const utils = require('./utils.js');
3
4   const stringedEvent = event.Records[0].body.toString('utf-8');
5   const eventParsed = JSON.parse(stringedEvent);
6   const stringedBody = JSON.stringify(eventParsed);
7   const bodyParsed = JSON.parse(stringedBody);
8   const check = bodyParsed.body;
9
10  const checkNameParams = {
11    TableName: 'role',
12    ExpressionAttributeNames: {
13      "#rolename": "name"
14    },
15    ProjectionExpression: "#rolename",
16    FilterExpression: "#rolename = :checkname",
17    ExpressionAttributeValues: {

```

```
18         ":checkname": check.name
19     }
20
21     const nameAlreadyExists = await utils.asyncCheckScanDB(checkNameParams);
22
23     if ((check.roleId == "" || check.name == "" || check.desc == "") || ↵
24         nameAlreadyExists)
25         callback(null, "Name already exists or empty attributes");
26     else{
27         utils.storeEvent("role", "executeCreateRoleQueue", bodyParsed.body);
28         callback(null, "Role event stored");
29     }
30 };
```


2.1.1.3 Mediator

This function catch the *DynamoDB* event and triggered when a change occurs in a certain table.

You have to be careful that the *DynamoDB* events are mapped using the char type attribute value. This is an example:

```
{
  "eventId": {
    "S": "bf6dffb9-72d8-ae5f-21fa-56dd6a26d572"
  },
  "payload": {
    "M": {
      "auth": {
        "S": "{\\n\\t\\\"Authorizations\\\": [\\n\\t\\t\\\"FullAccess\\\" ] }"
      },
      "roleId": {
        "S": "63c471c1-e5c7-09d0-ea8a-f20b27a0575c"
      },
      "name": {
        "S": "Admin"
      },
      "desc": {
        "S": "Full access to all resources"
      }
    }
  },
  "aggregate": {
    "S": "role"
  },
  "executionQueue": {
    "S": "executeCreateRoleQueue"
  },
  "timestamp": {
    "N": "1563358792450"
  }
}
```

Fig. 2: DynamoDB event object

You can use the *"AWS.DynamoDB.Converter"* module to parse a DynamoDB event object.

```

{
  "eventId": "bf6dffb9-72d8-ae5f-21fa-56dd6a26d572",
  "payload": {
    "auth": "{\n\t\"Authorizations\": [\n\t\t\"FullAccess\" ] }",
    "roleId": "63c471c1-e5c7-09d0-ea8a-f20b27a0575c",
    "name": "Admin",
    "desc": "Full access to all resources"
  },
  "aggregate": "role",
  "executionQueue": "executeCreateRoleQueue",
  "timestamp": 1563358792450
}

```

Fig. 3: Parsed DynamoDB event object

After that, the mediator retrieves the *executionQueue* parameter from the event object, the payload event passed with the *MessageBody* and then sends the execution message to corresponding SQS queue.

2.1.1.4 OperationAggregate

These functions execute a single operation using the event payload and triggered when new event arrives to the corresponding execution queue.

Example:

```

1 module.exports.createRole = async (event, context, callback) => {
2   const AWS = require('aws-sdk');
3   const dynamoDb = new AWS.DynamoDB.DocumentClient();
4
5   const stringedBody = event.Records[0].body.toString('utf-8');
6   const parsedBody = JSON.parse(stringedBody);
7
8   const params = {
9     TableName: 'role',
10    Item: parsedBody
11  };
12
13  await dynamoDb.put(params, (err, data) => {
14    if (err) {
15      console.log(err);
16      callback(null, err);
17    }
18    else
19      callback(null, "Role created");
20  }).promise();
21 };

```

2.1.1.5 Recovery

This function allows you to restore the system state starting from a given timestamp by replaying all the events into the *eventStore* table which were stored after that time.

Is important to re-execute the events one by one and in the correct order(from the oldest one); for this purpose the recovery function implements a sorting algorithm that retrieves an array of events and sorts them by timestamp, and an *async* function which sends every event to the corresponding execution queue.

2.1.2 Read mode Lambda functions

2.1.2.1 ReadOperationAggregate

These functions works in the read side of the architecture; they query the database to retrieve the informations needed without passing throw a SQS queue or a mediator. This kind of events aren't stored into the *eventStore* because they don't change the current system state; this Lambda functions fetch the URL of a GET endpoint using a GET request and listen for a response result.

```
fetch("https://domain/resourceAPIpath" [+ queryStringParam] {
  method: "GET",
  headers: {
    'Content-Type': 'application/json'
  }
});
```

Fig. 4: Fetch GET API

Example:

```
1 module.exports.getAllRoles = (event, context, callback) => {
2   const AWS = require('aws-sdk');
3   const dynamoDb = new AWS.DynamoDB.DocumentClient();
4
5   const params = {
6     TableName: 'role',
7     ExpressionAttributeNames: {
8       "#rolename": "name"
9     },
10    ProjectionExpression: "#rolename"
11  };
12
13  dynamoDb.scan(params, (err, data) => {
14    const stringifiedData = JSON.stringify(data);
15    if (err) {
16      console.log(err);
17      callback(null, err);
18    }
19  });
```

```
19     else{
20         if (data.Count == 0){
21             console.log("Role not found");
22             callback(null, "Role not found");
23         }
24         else{
25             const response = {
26                 statusCode: 200,
27                 headers: {
28                     'Content-Type': 'application/json',
29                     'Access-Control-Allow-Origin': '*',
30                     'Access-Control-Allow-Credentials': true
31                 },
32                 body: stringedData
33             };
34             callback(null, response);
35         }
36     }
37 });
38 };
```

2.1.3 CloudWatch Logs

Amazon CloudWatch is a monitoring and management service built for developers, system operators, site reliability engineers and IT managers. CloudWatch provides you with data and actionable insights to monitor your applications, understand and respond to system-wide performance changes, optimize resource utilization, and get a unified view of operational health. CloudWatch collects monitoring and operational data in the form of logs, metrics, and events.

This tool can be very helpful to debug your Lambda functions and to understand what happens to your system.

2.2 DynamoDB

Amazon DynamoDB is a fully managed proprietary NoSQL database service that supports key-value and document data structures and is offered by Amazon Web Services. In this project the DynamoDB instance is used for two purpose: it is used to store events and to keep updated the aggregates views.

The tables are the following:

- **eventStore**: to keep track of the occurred events;
- **user**: to store users info;
- **role**: to store roles info;
- **authorization**: to store authorizations info;
- **group**: to store groups info.

2.3 API Gateway

Amazon API Gateway is a fully managed service that makes it easy for developers to create, publish, maintain, monitor, and secure API at any scale. You can create REST and WebSocket API that act as a “front door” for applications to access data, business logic, or functionality from your backend services.

In this project, API endpoints starting the execution flow of a write mode function through a POST request, pushing a new event into the corresponding queue; in fact you have to provide an API endpoint for each function. On the other side, to query the database you have to reach the endpoint through a GET request to get a result response.

2.4 Simple Queue Service

Simple Queue Service(SQS) is a distributed message queuing service introduced by Amazon. It supports programmatic sending of messages via web service applications as a way to communicate over the Internet. SQS is intended to provide a highly scalable hosted message queue that resolves issues arising from the common producer-consumer problem or connectivity between producer and consumer.

In this project, for each function you have to use two queues:

- **OperationAggregateQueue:** this queues receive new events from an API endpoint and trigger the corresponding *commandOperationAggregate* function;
- **ExecuteOperationAggregateQueue:** this queues receive new events from the *mediator* function and trigger the corresponding *operationAggregate* function.

3 Serverless Framework

3.1 Description

The Serverless Framework is a free and open-source web framework written using Node.js. Serverless is the first framework that was originally developed for building applications exclusively on AWS Lambda, a serverless computing platform provided by Amazon as a part of the Amazon Web Services. Currently, applications developed with Serverless can be deployed to other function as a service providers, including Microsoft Azure with Azure Functions, IBM Bluemix with IBM Cloud Functions based on Apache OpenWhisk, Google Cloud using Google Cloud Functions, Oracle Cloud using Oracle Fn, Kubeless based on Kubernetes, Spotinst and Webtask by Auth0.

A Serverless app can simply be a couple of lambda functions to accomplish some tasks, or an entire back-end composed of hundreds of lambda functions. Serverless currently supports Node.js and Python runtime.

3.2 Serverless.yml

One advantage of using this tool is the capability to deploy every time an entire serverless system based on cloud providers, in this case AWS. You can define a *serverless.yml* configuration file which contains all the informations about your service. You don't need to manually create the resources you need in a project, like database tables, queues, API endpoints and Lambda functions. You just have to write the code and then deploying your app.

Example of *serverless.yml* configuration for User Management:

- Name of your services and list of plugins: the *split-stack* plugin is useful when you reach the limit of 200 resources to deploy because it automatically split your resources using nested stacks;

```
service: serverless-user-management

plugins:
  - serverless-plugin-split-stacks
```

Fig. 5: Plugins - *serverless.yml*

- About your cloud provider: you can also define your IAM role authorization policies;

```
provider:
  name: aws
  runtime: nodejs10.x
  region: eu-central-1
  stage: dev
  iamRoleStatements:
    - Effect: "Allow"
      Resource: "*"
      Action:
        - "dynamodb:*"
        - "sqs:*"
        - "lambda:*"
        - "cloudwatch:*"
        - "apigateway:*
```

Fig. 6: Provider info - serverless.yml

- DynamoDB tables parameters: you have to define the name of the table, name and type of the key and the provisioned throughput;

```
UsersDynamoDBTable:
  Type: 'AWS::DynamoDB::Table'
  Properties:
    AttributeDefinitions:
      - AttributeName: userId
        AttributeType: S
    KeySchema:
      - AttributeName: userId
        KeyType: HASH
    ProvisionedThroughput:
      ReadCapacityUnits: 5
      WriteCapacityUnits: 5
    TableName: user
```

Fig. 7: DynamoDB tables - serverless.yml

- Lambda function and its handler: you can also define which events trigger that lambda. In this case the trigger event is when new object arrives into the specified queue;

```
commandUpdateUser:
  handler: handler.commandUpdateUser
  timeout: 10
  events:
    - sqs:
        arn: arn:aws:sqs:eu-central-1:582373673306:updateUserQueue
```

Fig. 8: Lambda triggered by SQS event - serverless.yml

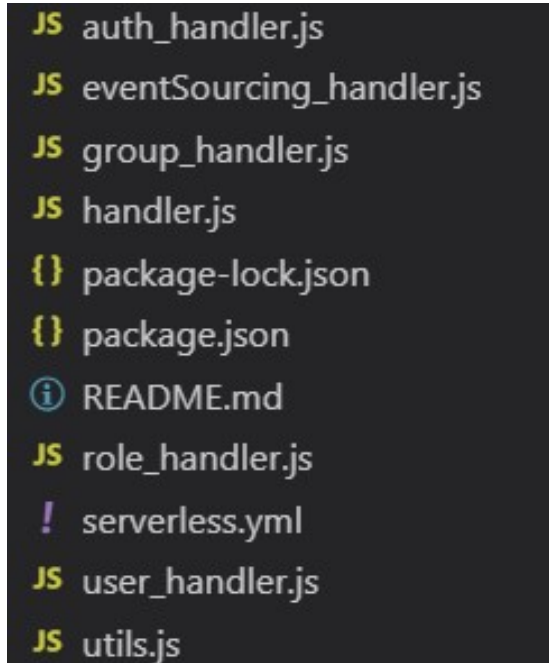
- You can define an API gateway endpoint which trigger the corresponding function: you can also use a custom request template.

```
readUser:
  handler: handler.readUser
  events:
    - http:
        path: /readUser
        method: get
        cors: true
        integration: lambda
        request:
          template:
            application/json: '{ "userId": "$input.params("userId")" }'
```

Fig. 9: Lambda triggered by GET endpoint - serverless.yml

3.3 Project's root

The main project's handler is composed of the other handlers, one per aggregate and another one to handle event sourcing functions. In this way the responsibilities are restricted to every type of aggregate.



```
JS auth_handler.js
JS eventSourcing_handler.js
JS group_handler.js
JS handler.js
{} package-lock.json
{} package.json
ⓘ README.md
JS role_handler.js
! serverless.yml
JS user_handler.js
JS utils.js
```

Fig. 10: Project's root

4 CQRS

Command Query Responsibility Segregation (CQRS) is an architectural pattern which separates the responsibility for modifying data (Command) from reading them (Query). The use of two different models for writing and reading operations, in scope of CQRS, allows instead to design and optimize each model for its responsibilities. In addition to this, the use of distinct models also allows the selection of the most appropriate technologies. As soon as the reading and writing models are separated, the infrastructure could easily scale to best fit the needs. It often happens that the number of writings in a system is much lower than the readings. Obviously the two models must be synchronized to ensure that the read information are consistent with the written ones.

The justification for CQRS is that in complex domains, a single model to handle both reads and writes gets too complicated, and we can simplify by separating the models.

The change that CQRS introduces is to split that conceptual model into separate models for update and display, which it refers to as Command and Query.

CQRS fits well with event-based programming models. It's common to see CQRS system split into separate services communicating with Event Collaboration. This allows these services to easily take advantage of Event Sourcing.

4.1 Architecture overview

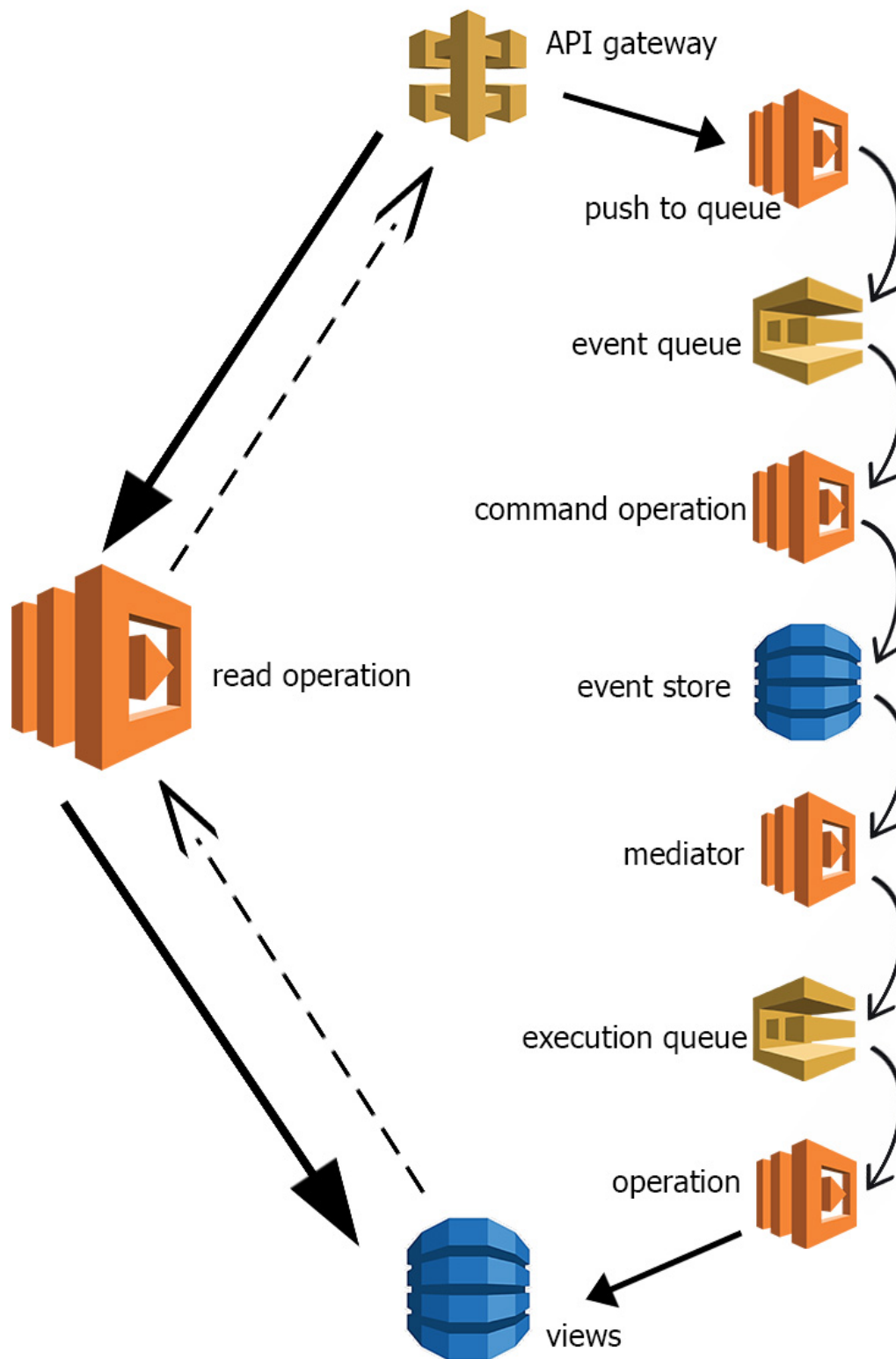


Fig. 11: Architecture overview

As you can see from the architecture overview, the application is separated into two models: write model(right side) and read model(left side).

The choice to apply the CQRS pattern due to separate the responsibilities of writing and reading, but also because read side events aren't stored into the event store table because they don't change the system's state but just retrieve information from it.

4.2 Write model

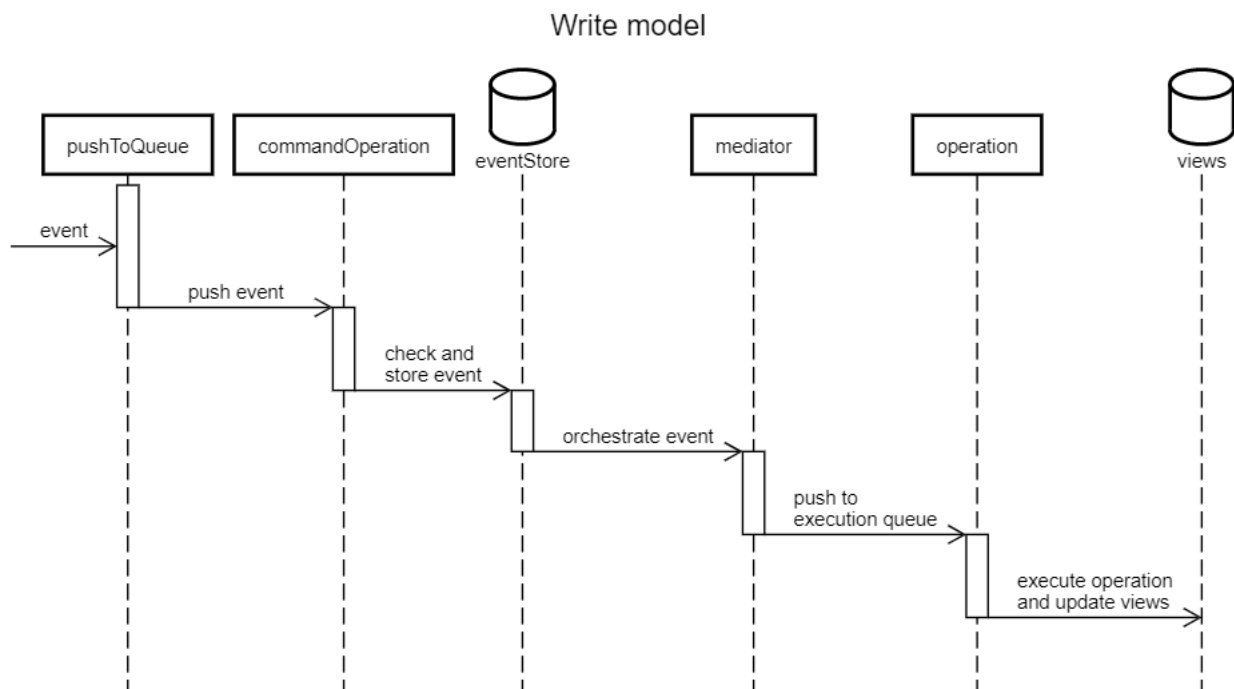


Fig. 12: Write model sequence diagram

4.3 Read model

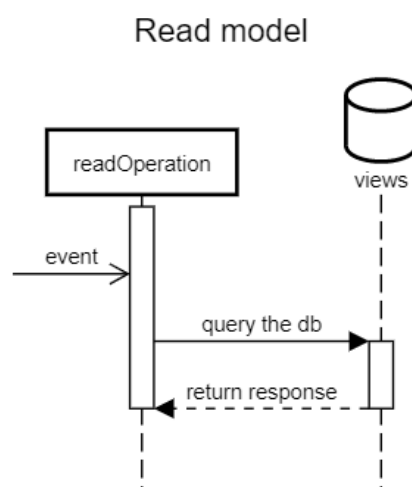


Fig. 13: Read model sequence diagram

5 User Management

5.1 Aggregates

5.1.1 User

- **UserId**: this attribute must be unique and refers to the *user_id* key in the corresponding *Auth0* users table(without first 6 characters 'Auth0|');
- **FirstName**: user's first name;
- **LastName**: user's last name;
- **Date**: user's birth date;
- **Email**: this attribute must be unique and respect the right format;
- External links:
 - **Role**: user's role;
 - **Group**: user's belonging group.

5.1.2 Role

- **RoleId**: this attribute is generated with a UUID function;
- **Name**: this attribute must be unique;
- **Desc**: role's description;
- External links:
 - **Auth**: JSON array of role's authorizations.

5.1.3 Authorization

- **AuthId**: this attribute is generated with a UUID function;
- **Name**: this attribute must be unique;
- **Desc**: auth's description.

5.1.4 Group

- **GroupId**: this attribute is generated with a UUID function;
- **Name**: this attribute must be unique;
- **Desc**: group's description.

5.2 Admin side

5.2.1 Authentication

The authentication feature is provided by a third part provider, Auth0. In the admin side only admin users can log in; the user must be in the corresponding application's database on Auth0.

5.2.2 Use cases

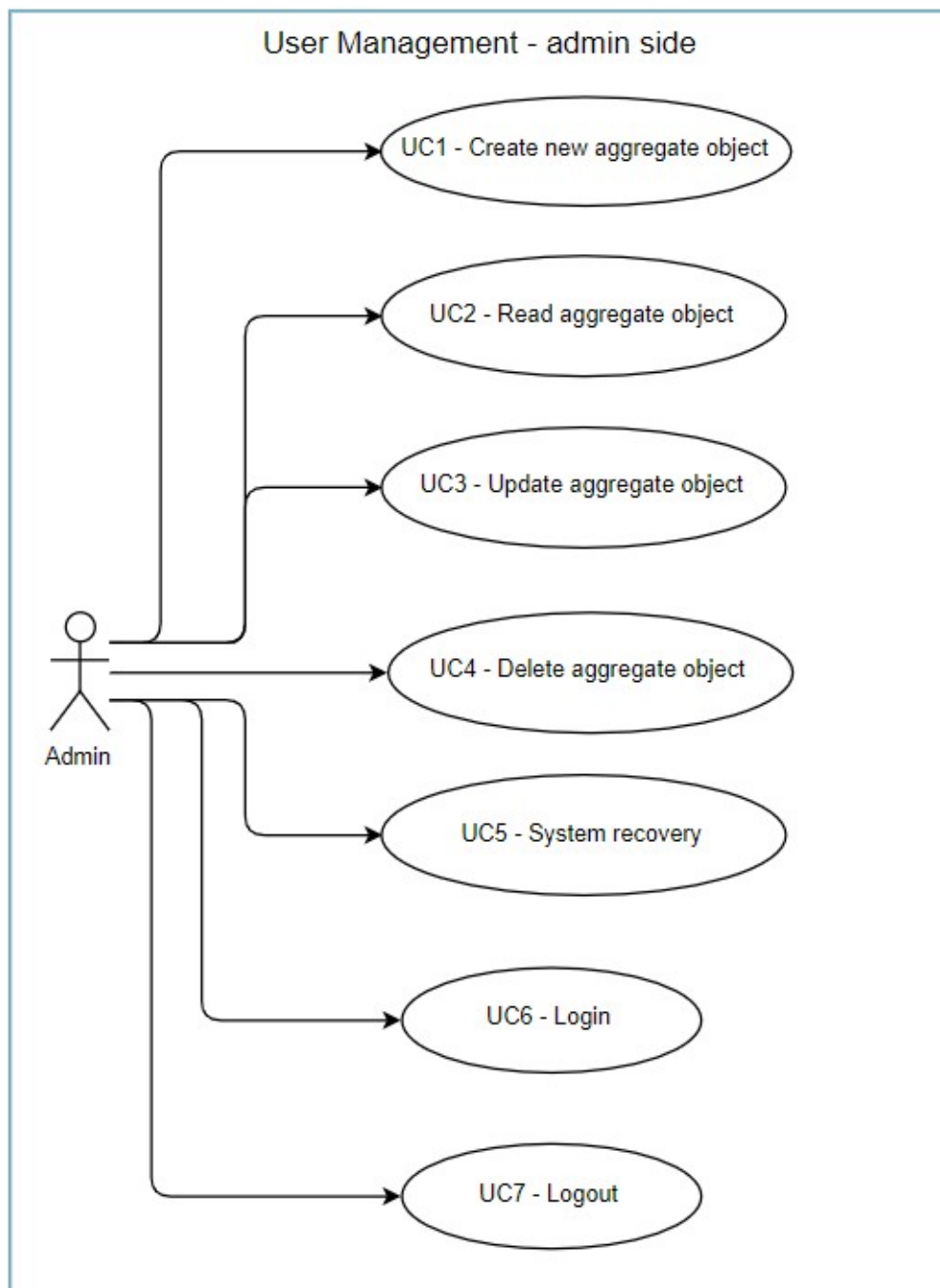


Fig. 14: Admin use cases

- UC1: the admin creates a new object based on the type of aggregate;
- UC2: the admin read the information about an object based on the type of aggregate;
- UC3: the admin updates an object based on the type of aggregate;
- UC4: the admin deletes an object;
- UC5: the admin recovers the system state from the chosen timestamp re-running of the event occurred after that;
- UC6: the admin log in into the application using the Auth0 portal;
- UC7: the admin log out from the application.

5.3 User side

5.3.1 Authentication

The authentication feature si provided by a third part provider, Auth0. In the user side only users can log in; it can sign in with email and password or using Google's authentication or just log in using its credentials.

5.3.2 Use cases

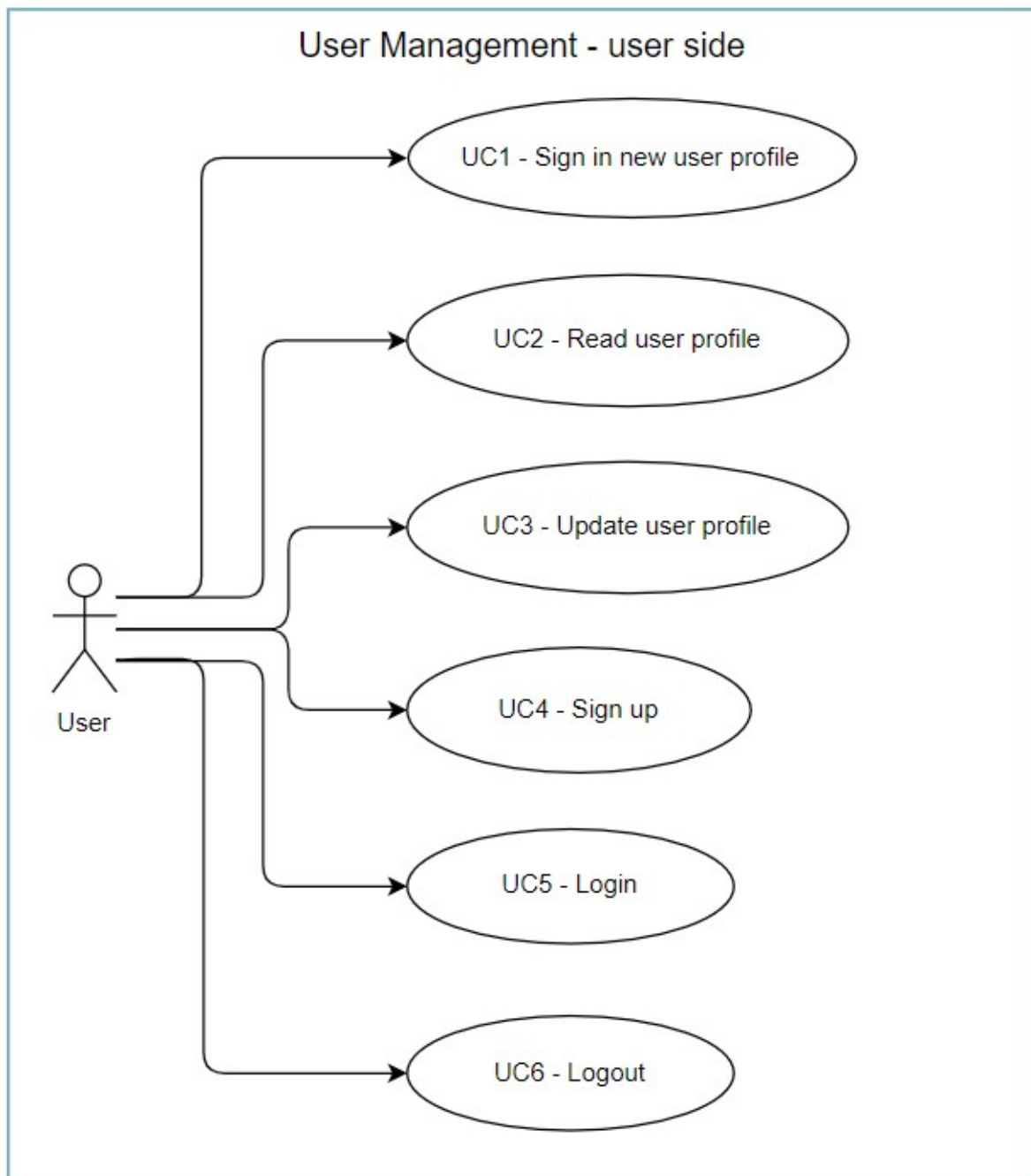


Fig. 15: User use cases

- UC1: the user fills the form to sign in into the application the first time logs in;
- UC2: the user read his profile;
- UC3: the user updates his profile;
- UC4: the user sign up to Auth0 authentication portal using email and password or Google's account;

- UC5: the user log in to Auth0 authentication portal using email and password or Google's account;
- UC6: the user log out from the application.

6 Event Sourcing

"Event Sourcing ensures that all changes to application are stored as a sequence of events."

You can query these events, use the event log to reconstruct past states and adjust the state to cope with retroactive changes.

The core idea of event sourcing is that whenever we make a change to the state of a system, we record that state change as an event, and we can confidently rebuild the system state by reprocessing the events at any time in the future.

When working with an event log can be useful to build snapshots of the working copy so that you don't have to process all the events when you rebuild the system or every time you need to query the database. For this reason when a new event is stored triggers another function to update the views.

Event sourcing can be used to:

- **Complete Rebuild:** you can discard the application state completely and rebuild it by re-running the events from the event store on an empty application.
- **Temporal Query:** you can determine the application state at any point in time. This can be used considering multiple time-lines (like branching in a VCS).
- **Event Replay:** if you find a past event was incorrect, you can replaying from then with the new event. The same technique can handle events received in the wrong sequence with systems that communicate with asynchronous messaging.

6.1 Event store

Event store is a database's table that contains all the events occurred from the begin. When using Event sourcing the event store becomes the principal source of truth and the system state is completely derived from it.

6.1.1 Event object

Each event stored in the table is a JSON object and respect this format:

- **EventId:** is a UUID;
- **Payload:** JSON object which contains all the information about the event;
- **Aggregate:** type of the aggregate that the event refers to;
- **ExecutionQueue:** name of the execution queue used to execute the event;
- **Timestamp:** number to know the execution order of the events.

```
aggregate String : user
eventId String : f7731a40-f634-354a-ac99-351074d83b6d
executionQueue String : executeCreateUserQueue
▶ payload Map {7}
timestamp Number : 1563543008991
```

Fig. 16: DynamoDB eventStore example item

7 Extension points

7.1 New aggregates

To create a new type of aggregate you have to:

- Define a new DynamoDB table in the *serverless.yml*;
- Create an *operationQueue* and an *executionQueue* with SQS;
- Define all the Lambda functions you need in the *serverless.yml*;
- Define the trigger events in the *serverless.yml*;
- Create a handler file named *"aggregate_handler.js"* in the project's root which contains the Lambda's handlers that refer the same aggregate type;
- Add the new handlers file in the *module.exports* of the main handler;
- Deploy the serverless application.

7.2 New write functions

To add a new write function you have to:

- Define the function's name and the corresponding handler's name in the *serverless.yml*;
- Define trigger events in the *serverless.yml*;
- Create an *operationAggregateQueue* and an *executeOperationAggregateQueue*;
- In the *"aggregate_handler.js"* you must write:
 - a *pushOperationAggregateToSQS* function to push the event in the corresponding queue;
 - a *commandOperationAggregate* function to check if the event is valid and store it into the *eventStore*;
 - an *operationAggregate* function to execute the event and update the view.
- Deploy the serverless application.

7.3 New read functions

To add a new read function you have to:

- Define the function's name and the corresponding handler's name in the *serverless.yml*;
- Define the corresponding get API endpoint in the *serverless.yml*;
- Write the *readOperation* function to query the database;
- Deploy the serverless application.

8 Setup