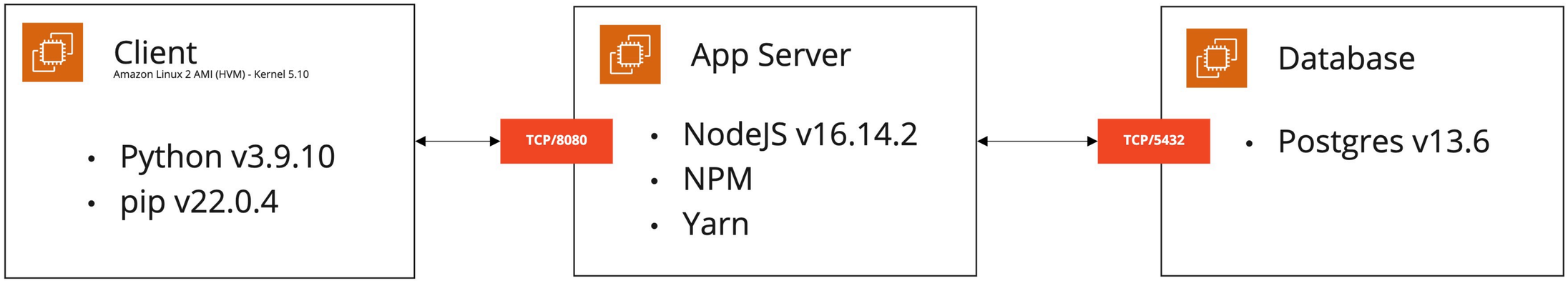
Data Sync and Cost Optimization

2110524 - Cloud Computing Tech (2022/2) - Midterm Project

Panya Batphantana | March 2022

System Tier and Tech Stack



Data Storage strategies

Inbound data

Create:
- Insert with active flag to true

Update:
- Overwrite the attributes: author, message, likes

Delete:
- Overwrite the active flag to false

data_enhanced	
uuid: char(36)	PK
author: varchar(64)	
message: varchar(1024)	
likes: bigint *default(0)*	
active: bool *default(true)*	IDX^
timestamp_key: char(25)	IDXv

Update:
- Insert the entity id with change type = 'U' and latest state of data after the update.

Delete:
- Insert the entity id with change type = 'D', left state unwritten.

entity_memento	
id: char(25)	PK, IDXv
entity_id: char(36)	
change_type: char(1)	IDX
state: jsonb *default(0)*	

Add-On Techniques in Data Storage

Using CUID as a sorting key
- CUID is the unique id algorithms composed from unix timestamp and machine fingerprint
- Suffixing with unix timestamp having advantage in data which came in time-series manner
- Ref: https://github.com/ericelliott/cuid

UPDATE = DELETE under the hood
- In Postgres, data is UPDATE or DELETE with the new tuple of value.
- The old tuple value was left un-referred until postgres start the VACUUM process which merely like a Garbage Collection in JVM
- DELETE operation might catalyst the VACUUM as the deleting data is absolutely unrelated.

Event-Driven / Memento
- To enable the read from main table, entity, going in single direction, memento record the movement of existing data in time manner.
- Instead of traverse back and forth when read the main table, client may use memento as a historical / transactional record of particular row.

Server Architectural Design

App Server

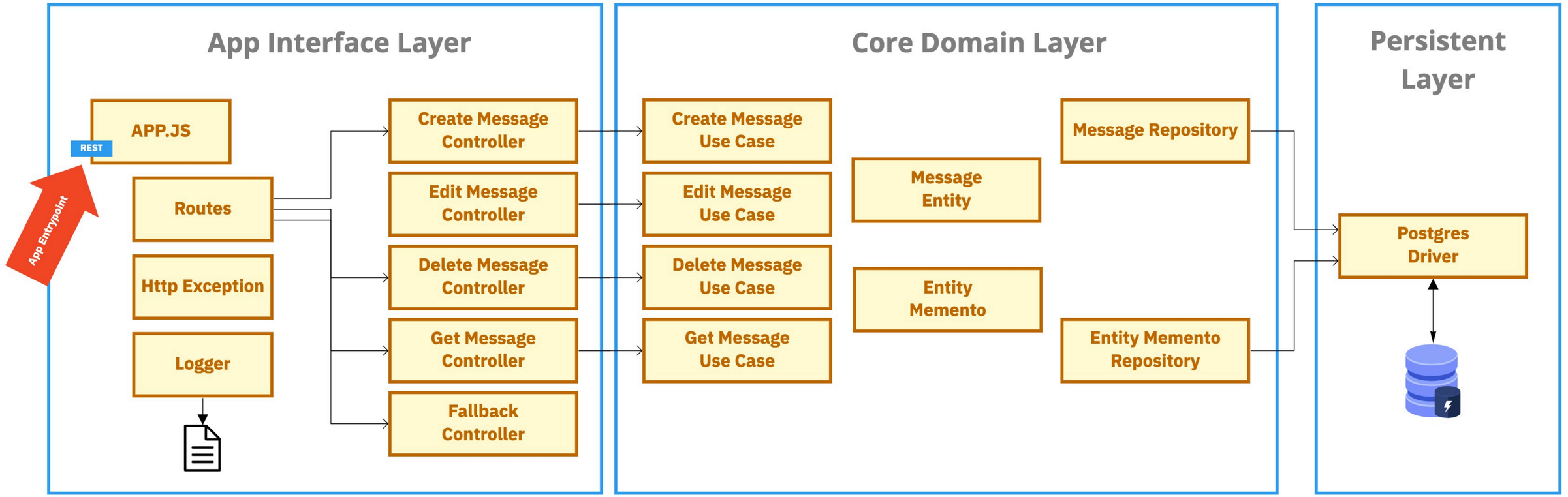
- NodeJS v16.14.2
- NPM
- Yarn

Implemented Language:

- JavaScript (CommonJS)

Application dependencies

- Express.JS - HTTP API framework
- PG - Postgres Driver for NodeJS
- Winston - Logger library



Response of Get Message API

GET /api/messages


```
{
    "c": [
        [
            "912a2576-a171-4d3f-9170-fd2ec64559a8",
            "Rick_Astley",
            "Never gonna give you up never gonna let you down",
            1163061387
        ],
        [
            "81a5bf0f-42c9-4078-91af-8a5b791e9823",
            "grizzledThrush5",

"6419f344bf89357e0351dc92d2017969cd820963a13a0de0c4bbfe2ca85a2dd3fb6bdf248acee05d87586020fafd2322d832cb3c38ed61e8759c9a
7af014b48b0a68e7f6eba4fce647a2416b93920d504502df1de2eef4a5719d78eb327386b441442ef3b808109d94a6aa54b878f3369b17551f3a2a7
02fc63763522bfca877bd38e448e19ecfd553aebd74b4f3595321c2b0dcb9355c02c5711a4bba665f6a2037ddc479d3f87c09ce6b54f667be748170
c13e53ee3acc560faa3acf1555fef14521474a75ef7f46008ca13cb9ded689f521aec86d6f0f11355045bdbd67d7bf44ee52aaaf423264ac2229464
2a031e3462c4741c85b01d6e504ba32a66f8a19977e4288a7ae1b43331e27d4787ed98691c1647da9441aa389c6c50031f93daf4423e5355e314dd7
07738837fcdb0f",
            321099
        ]
    ],
    "u": [
        [
            "6d859661-3b20-47ad-be95-bf9c71115d74",
            "Olivia Price",
            "test-updated-quantifying the program won't do anything, we need to compress the neural SSL application!",
            222
        ]
    ],
    "d": "036ef391-31ef-4810-8192-f7e9a911a40f,afd07954-d14f-4248-8093-807969090f40"
}
```


KEY	VALUE
X-Powered-By	Express
x-create-cursor	cl10zbwyc0001lp9m6m5m0ltm
x-delete-cursor	cl12krnb40009jv9mcv4mb6ag
x-update-cursor	cl12mtta900019n9mg6pk37c3
Content-Type	application/json; charset=utf-8
Content-Length	1053
ETag	W/"41d-HnCs4TJZbx0MGKpF99m/Zm+/O7s"
Date	Tue, 22 Mar 2022 21:13:37 GMT
Connection	keep-alive
Keep-Alive	timeout=5

Noted: API returns 3 cursor in response header with *x-*-cursor* for next request pagination request.

Query Parameters of Get Message API

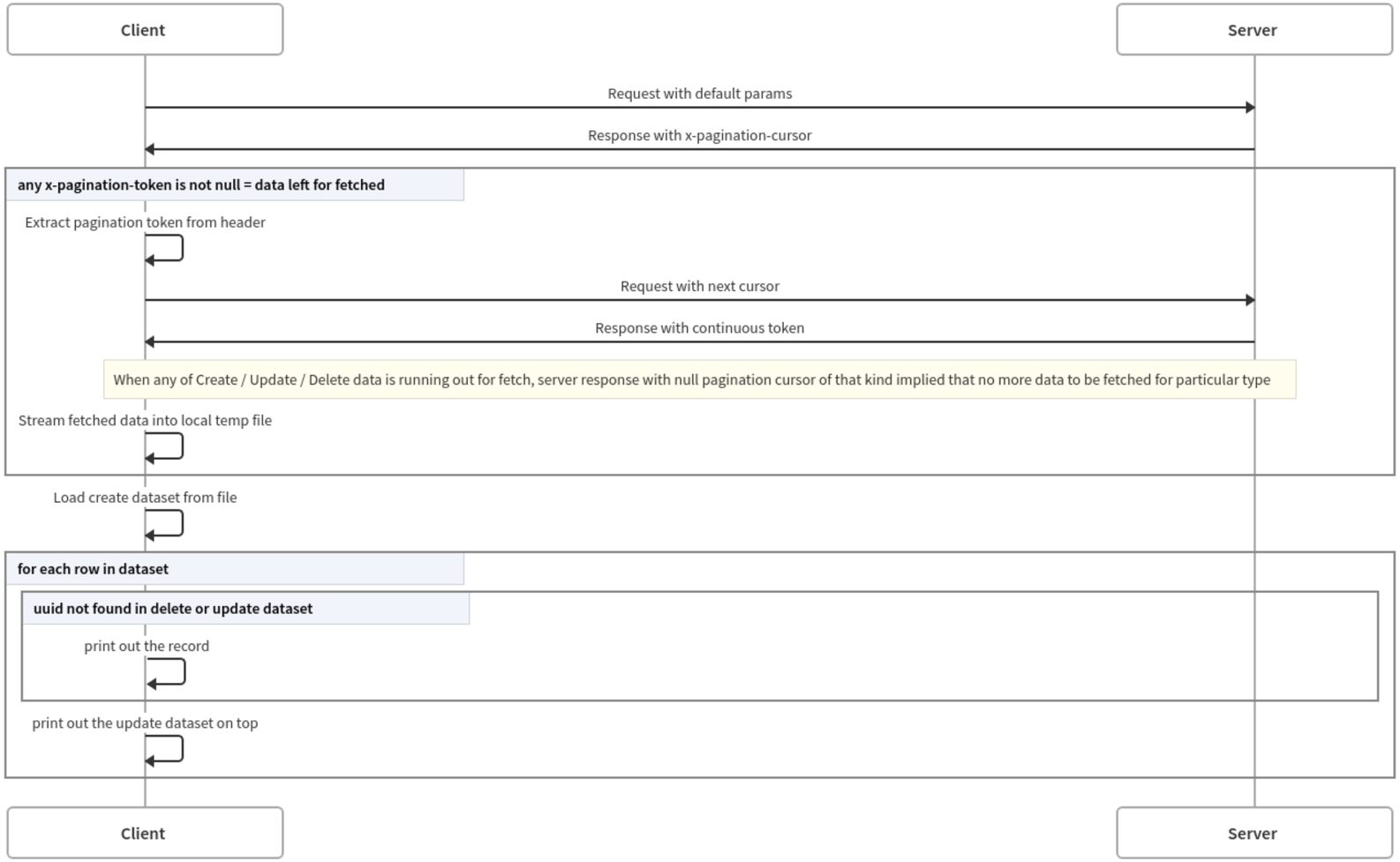
GET /api/messages

	KEY	VALUE	DESCRIPTION
☑	limit	2	Limit of result set in create and delete dataset (Default: 10000)
☑	update_cursor	cl12kvwdy1i0djv9mdlha1fwl	Cursor acquired from x-update-cursor in previous request [Optional]
☑	delete_cursor	cl12kvwdu1i0bjv9m2v8y9i7l	Cursor acquired from x-delete-cursor in previous request [Optional]
☑	create_cursor	cl12kvwe31i0fjv9maow6bi79	Cursor acquired from x-create-cursor in previous request [Optional]

Noted: *update_cursor*, *create_cursor*, *delete_cursor* are optional parameters which can be omitted to fetch data from the beginning of table

Client Flow

Sequential interaction between client and server



HTTP Request with Python

3rd-Party Library requests had hidden overhead!

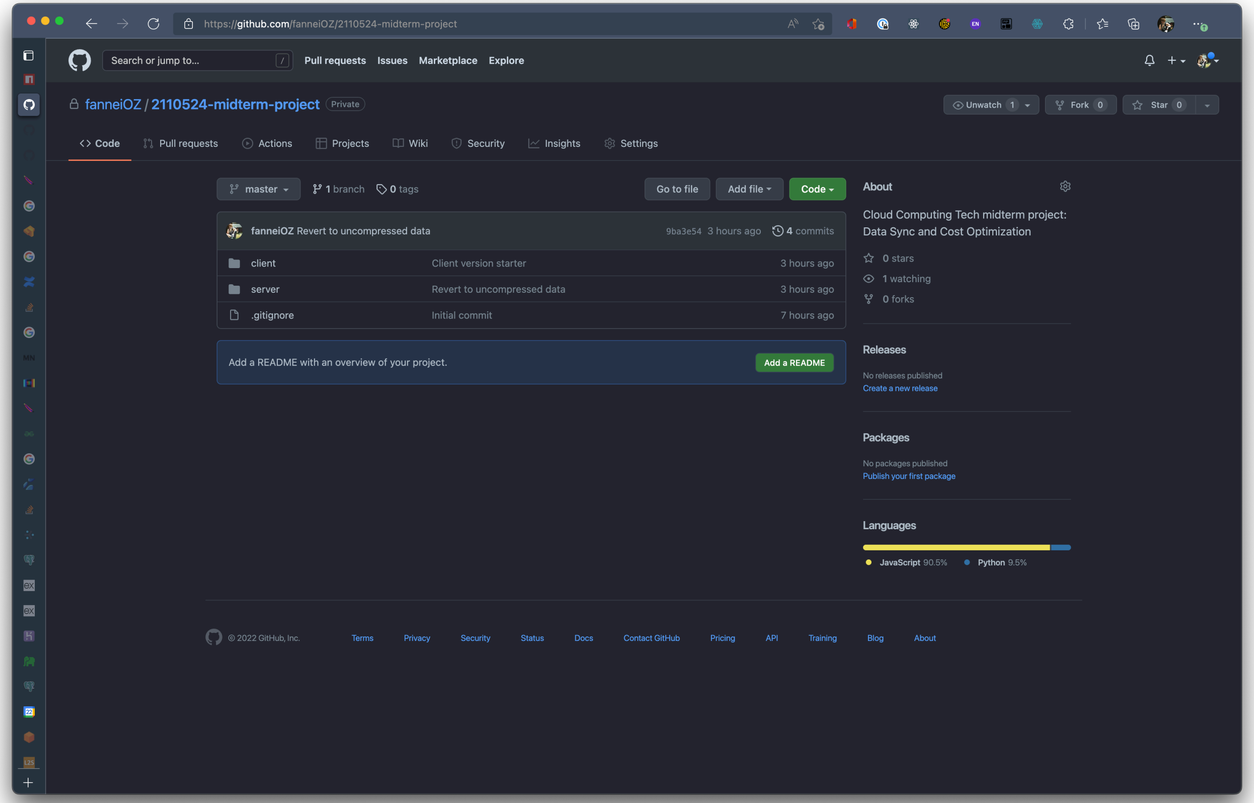
```python
import requests
import time
import http.client
import urllib.request

headers = {'Accept': 'application/json', 'User-Agent': 'python/3.10'}

tic = time.perf_counter()

# Using http.client <- Support HTTP natively, HTTPS but compiled with SSL
conn = http.client.HTTPConnection('www.colourlovers.com')
conn.request(method='GET', url='/api/color/6B4106?format=json', headers=headers)
decoded_res = conn.getresponse().read().decode('utf-8')
conn.close()
print(F'Low Lv API#1 - http.client -> Elapsed time: {time.perf_counter() - tic:0.4f}s')
# print(decoded_res)

tic = time.perf_counter()

# Using urllib.request
req = urllib.request.Request(url='http://www.colourlovers.com/api/color/6B4106?format=json', method='GET', headers=headers)
decoded_res = urllib.request.urlopen(req).read().decode('utf-8')
print(F'Low Lv API#2 - urllib.request -> Elapsed time: {time.perf_counter() - tic:0.4f}s')
# print(decoded_res)

tic = time.perf_counter()
# Using requests lib
res = requests.request(method='GET', url='http://colourlovers.com/api/color/6B4106', params={'format': 'json'}, headers=headers)
decoded_res = res.text
print(F'High Lv API - requests lib -> Elapsed time: {time.perf_counter() - tic:0.4f}s')
# print(decoded_res)

```

```
Run: asyncio ×
/Users/funnyfeb/.pyenv/versions/3.9.7/bin/python /Users/funnyfeb/research/repos/2110524-cloud-comp-tech/05-s3/asyncio.py
Low Lv API#1 - http.client -> Elapsed time: 0.7086s
Low Lv API#2 - urllib.request -> Elapsed time: 0.6459s
High Lv API - requests lib -> Elapsed time: 1.2526s
```

Make HTTP request with requests lib is easier as a high-level API but it take almost 50% additional runtime comparing to low-level API

Source Code

https://github.com/fanneiOZ/2110524-midterm-project



Q&A

Thank You