

به نام خدا

ندا زارعی - ۴۰۰۲۳۰۳۵

گزارش تمرین سوم برنامه نویسی وب

کرل ریکوست:

```
[{"id": "21", "message": "file upload initiated successfully"}]
→ ~ curl -X POST http://127.0.0.1:8000/api/v1/upload \
-H "Content-Type: application/json" \
-d '{
  "file_url": "https://cdn.pixabay.com/photo/2024/02/28/07/42/european-shorthair-8601492_640.jpg",
  "bucket_name": "file-bucket",
  "object_name": "kitty2.jpg"
}'
```

سمت producer:

```
kafka-ui | 2024-11-24 17:21:01,062 DEBUG [parallel-1] c.p.k.u.s.ClustersStatisticsScheduler: Start getting metrics for kafkaCluster: local
kafka-ui | 2024-11-24 17:21:01,076 DEBUG [parallel-6] c.p.k.u.s.ClustersStatisticsScheduler: Metrics updated for cluster: local
web-1 | {"time": "2024-11-24T17:21:05.57172809Z", "id": "", "remote_ip": "192.168.64.1", "host": "127.0.0.1:8000", "method": "POST", "uri": "/api/v1/up
load", "user_agent": "curl/7.81.0", "status": 202, "error": "", "latency": 1027270726, "latency_human": "1.027270726s", "bytes_in": 200, "bytes_out": 59}
kafka-ui | 2024-11-24 17:21:08,992 DEBUG [boundedElastic-1] c.p.k.u.e.RangePollingEmitter: Starting polling for ConsumerPosition(seekType=BEGIN
NING, topic=upload_file_topic, seekTo=null)
```

سمت consumer که یک presigned url میدهد:

```
y-minio-file.darkube.app/file-bucket/kitty.jpg?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=c4JrL9xJILzgIH7zRu4y%2F20241124%2Fus-east-1%2Fs3
%2Faws4_request&X-Amz-Date=20241124T172727Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=08ab75feaac32731ecb94b62418e02dfff6e0aed
2e0cc79a6084a9b4200e6d4
web-1 | 2024/11/24 17:27:27 successfully processed file kitty2.jpg. ETag: 1c08478abf509b7b0dcb32d625151d2b, Size: 43410, presigned URL: https://
my-minio-file.darkube.app/file-bucket/kitty2.jpg?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=c4JrL9xJILzgIH7zRu4y%2F20241124%2Fus-east-1%2F
s3%2Faws4_request&X-Amz-Date=20241124T172727Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=847a2d7854099bd4fe604bfc90bdb958c75b66a
e42091bc26d7a7f4607e27cca
```

نتیجه در کافکا localhost:8080:

18 0 11/24/2024, 20:51:05 {"id": "21", "file_url": "https://cdn.pixabay.com/photo/...

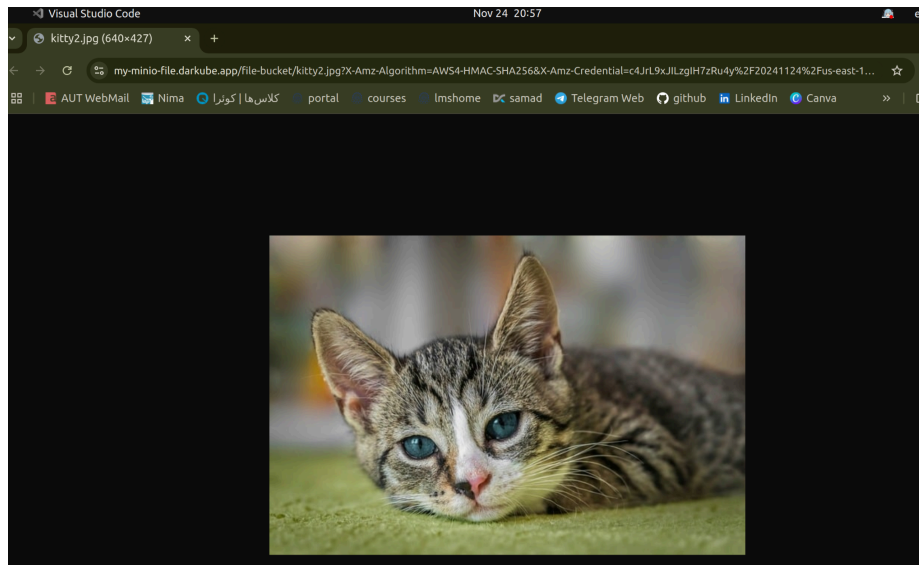
Key	Value	Headers	Timestamp	11/24/2024, 20:51:05
			Timestamp type:	CREATE_TIME
			Key Serde	Size: 0 Bytes
			Value Serde	String Size: 159 Bytes

```
{
  "id": "21",
  "file_url": "https://cdn.pixabay.com/photo/2024/02/28/07/42/european-shorthair-8601492_640.jpg",
  "bucket_name": "file-bucket",
  "object_name": "kitty2.jpg"
}
```

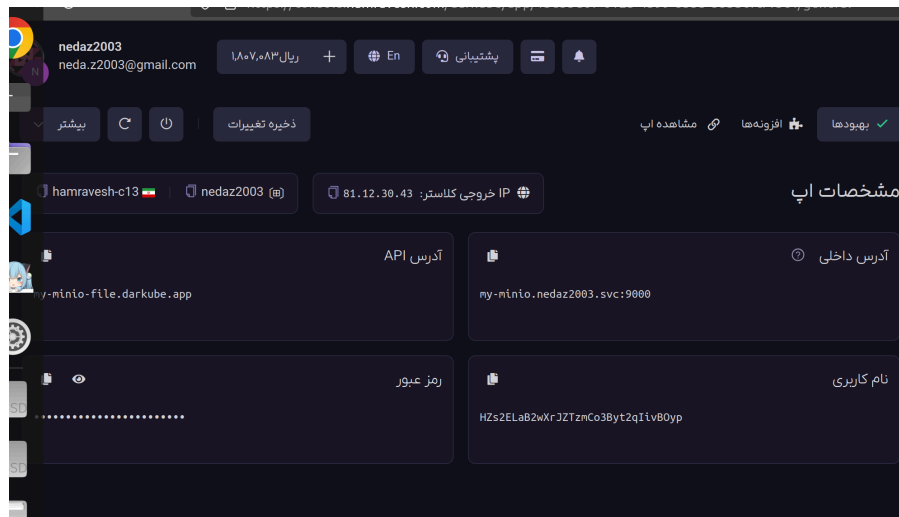
نتیجه در دیتابیس در تیل فایل:

id	file_url	bucket_name	object_name	etag
8	local://sonny.jpg	file_bucket	sonny.jpg	
16	https://www.aiseesoft.com/images/tutorial/jpg-to-url/jpg-to-url.jpg	file-bucket	sonny.jpg	6cd21411032f683b49fb866d52b
17	https://www.aiseesoft.com/images/tutorial/jpg-to-url/jpg-to-url.jpg	file-bucket	sonny.jpg	6cd21411032f683b49fb866d52b
18	https://cdn.pixabay.com/photo/2024/02/28/07/42/european-shorthair-8601492_640.jpg	file-bucket	cat.jpg	1c08478abf509b7b0dcb32d6251
19	https://cdn.pixabay.com/photo/2024/02/28/07/42/european-shorthair-8601492_640.jpg	file-bucket	kitten.jpg	1c08478abf509b7b0dcb32d6251
20	https://cdn.pixabay.com/photo/2024/02/28/07/42/european-shorthair-8601492_640.jpg	file-bucket	kitty.jpg	1c08478abf509b7b0dcb32d6251
21	https://cdn.pixabay.com/photo/2024/02/28/07/42/european-shorthair-8601492_640.jpg	file-bucket	kitty2.jpg	1c08478abf509b7b0dcb32d6251

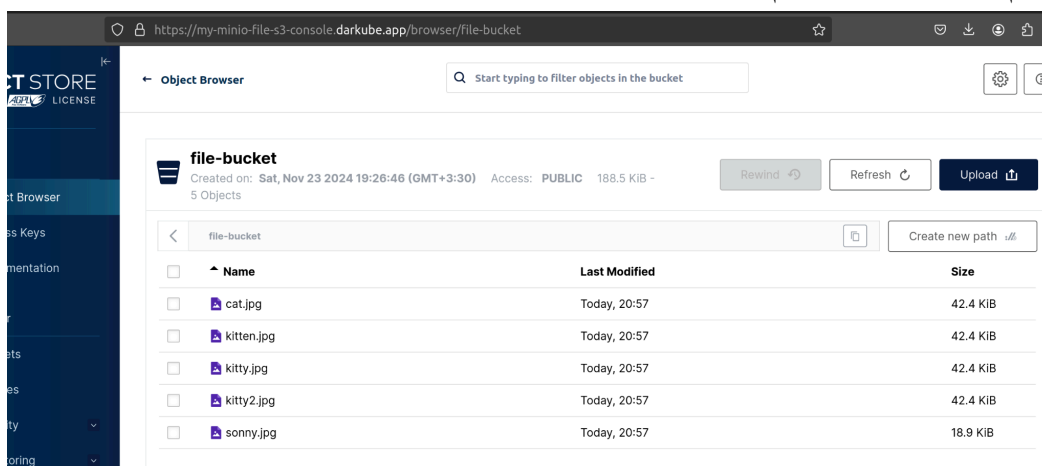
با کلیک کردن روی url فایل باز میشود:



برای کافکا و postgres از داکر استفاده شد و برای استفاده از minio از همروش استفاده کردم :



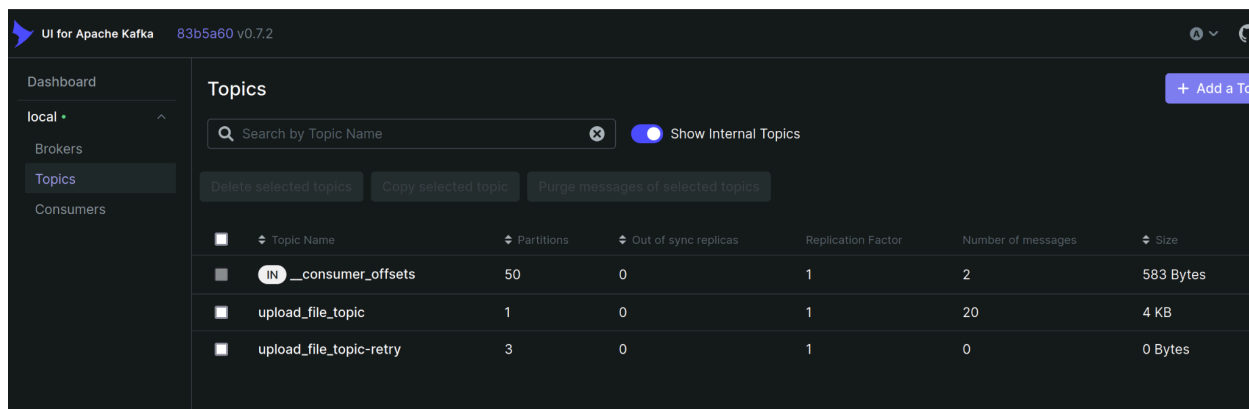
و اسم باکت هم file-bucket قرار دادم:



برای بخش امتیازی:

مکانیزم اصلی retry:

یک تاپیک کافکای جدید به نام `upload_file_topic-retry` اضافه کردم
وقتی آپلود یا دانلود فایل با خطا مواجه میشه، فایل به این تاپیک جدید فرستاده میشه
سیستم تا 3 بار سعی میکنه دوباره عملیات رو انجام بده
اگر بعد از 3 بار همچنان خطا داشت، فایل رو drop میکنه



Topic Name	Partitions	Out of sync replicas	Replication Factor	Number of messages	Size
__consumer_offsets	50	0	1	2	583 Bytes
upload_file_topic	1	0	1	20	4 KB
upload_file_topic-retry	3	0	1	0	0 Bytes

تغییرات در دیتابیس:

```
→ p2 git:(master) x psql -h localhost -p 5433 -U postgres -d simpleapi_database
Password for user postgres:
psql (17.0 (Ubuntu 17.0-1.pgdg22.04+1), server 13.17 (Debian 13.17-1.pgdg120+1))
Type "help" for help.

simpleapi database=# ALTER TABLE files
ADD COLUMN status VARCHAR(20) DEFAULT 'pending',
ADD COLUMN retry_count INTEGER DEFAULT 0,
ADD COLUMN last_error TEXT,
ADD COLUMN updated_at TIMESTAMP DEFAULT CURRENT_TIMESTAMP;
ALTER TABLE
simpleapi database=# select * from files;
simpleapi database=# CREATE INDEX idx_files_status ON files(status);
CREATE INDEX
simpleapi database=#
```

وضعیت هر

فایل رو ذخیره میکنیم (pending, completed, failed)

تعداد دفعات retry رو نگه میداریم

آخرین خطای رخ داده رو ذخیره میکنیم

ساختار کد:

دو consumer جداگانه داریم:

1. consumer اصلی که فایل‌های جدید رو پردازش میکنه
2. consumer مخصوص retry که فایل‌های خطا خورده رو مجدداً پردازش میکنه

فرآیند retry:

وقتی خطایی رخ میده

```
if err := h.processFile(ctx, fileUpload, 0); err != nil {  
    میفرستیم retry فایل رو به تاپیک  
    h.sendToRetryTopic(ctx, fileUpload, err, 0)  
}
```

retry در تاپیک

```
if retryFile.RetryCount >= maxRetries {  
    رسیدیم retry اگر به حداکثر تعداد  
    h.logFailedFile(ctx, retryFile)  
} else {  
    دوباره سعی میکنیم  
    h.processFile(ctx, retryFile.FileUpload, retryFile.RetryCount+1)  
}
```

مانیتورینگ و لاگ:

هر retry در دیتابیس ثبت میشه وضعیت نهایی هر فایل مشخصه خطاها با جزئیات لاگ میشن

تنظیمات کافکا:

kafka:

- KAFKA_CFG_AUTO_CREATE_TOPICS_ENABLE=true

- KAFKA_CFG_TOPIC_FILEFLOW-RETRY_RETENTION_MS=86400000 24 نگهداری پیامها تا

ساعت

وقتی فایل جدید میاد، اول به تاپیک اصلی میره اگر خطا بخوره، به تاپیک retry منتقل میشه

سیستم تا 3 بار تلاش میکنه نتیجه نهایی در دیتابیس ثبت میشه

این سیستم کمک میکنه که فایلها به خاطر خطاهای موقتی از دست نرن

وضعیت هر فایل قابل پیگیری باشه

بعد از تعداد مشخصی تلاش، سیستم فایل رو رها کنه

مشکلات سیستمی به راحتی قابل دیباگ باشن

```
→ ~ curl -X POST http://127.0.0.1:8000/api/v1/upload \
-H "Content-Type: application/json" \
-d '{
    "file_url": "https://tinyjpg.com/images/social/website.jpg",
    "bucket_name": "file-bucket",
    "object_name": "panda.jpg"
}'

{"id":"22","message":"file upload initiated successfully"}
→ ~
```

لاگ های consumer:

```
web-1 | 2024/11/24 0:22:58 successfully processed file kitty.jpg on attempt 1
web-1 | 2024/11/24 20:22:58 successfully processed file kitty.jpg on attempt 1
web-1 | 2024/11/24 20:23:01 successfully processed file kitty2.jpg on attempt 1
web-1 | 2024/11/24 20:23:02 successfully processed file panda.jpg on attempt 1
```

دیتابیس جدید:

```

c314012515a2b3551f0a430404 | 24900 | completed | 0 | minio upload error: the specified bucket does not ex
ist | 2024-11-24 19:43:25.097345
16 | https://www.aiseesoft.com/images/tutorial/jpg-to-url/jpg-to-url.jpg | file-bucket | sonny.jpg | 6c
d21411032f683b49fb866d52b75e96 | 19353 | completed | 0 |
| 2024-11-24 19:43:25.097345
17 | https://www.aiseesoft.com/images/tutorial/jpg-to-url/jpg-to-url.jpg | file-bucket | sonny.jpg | 6c
d21411032f683b49fb866d52b75e96 | 19353 | completed | 0 |
| 2024-11-24 19:43:25.097345
18 | https://cdn.pixabay.com/photo/2024/02/28/07/42/european-shorthair-8601492_640.jpg | file-bucket | cat.jpg | 1c
08478abf509b7b0dc32d625151d2b | 43410 | completed | 0 |
| 2024-11-24 19:43:25.097345
19 | https://cdn.pixabay.com/photo/2024/02/28/07/42/european-shorthair-8601492_640.jpg | file-bucket | kitten.jpg | 1c
08478abf509b7b0dc32d625151d2b | 43410 | completed | 0 |
| 2024-11-24 19:43:25.097345
:

```