HIGH PERFORMANCE DATA PROCESSING WITH PYTHON, KAFKA AND ELASTICSEARCH

EUROPYTHON ONLINE 2021

Harshit Prasad / harshitprasad.com
INTRODUCTION

- Software Engineer at **Grofers - India**.
- Bachelors in Electronics and Communication Engineering from the **LNM Institute of Information Technology**
- Past GSoC'16 - **FOSSASIA**
- Past GSoC'17 - **CERN**
- Open Source Contributor.
TALK OUTLINE

● Problem Statement
● Solution
● Workflow
● Code Walkthrough
PROBLEM STATEMENT

- Your own application
- Avg order rate X
- Peak hour order rate X * 1000
PROBLEM STATEMENT

A diagram showing the interaction between a client, a database, and two services. The client sends a request to service 1, which in turn sends a request to the database. The database responds, and service 1 responds back to the client with a response. Service 2 is updated based on a specific action, and sends an event back to service 1.
PRODUCER CONSUMER MODEL

producer

message queue

service 1

service 2

service 3

consumers
WORKFLOW

```
elasticsearch
  update

service
  request

kafka
  request

response

retail shopping app
```
WHAT KIND OF DATA ARE WE UPDATING?

```
{
  "_index" : "retail-catalogue",
  "_type" : "_doc",
  "_id" : "10",
  "_version" : 3,
  "_seq_no" : 4,
  "_primary_term" : 2,
  "found" : true,
  "_source" : {
    "product" : {
      "id" : 10,
      "name" : "Funny Farm House Ketchup",
      "category" : "Dips and Ketchups",
      "price" : 15.0,
      "stock" : 5
    },
    "update_ts" : 1625931472
  }
}
```
CODE WALKTHROUGH

- Python based implementation of Producer Consumer model.
- Walkthrough of Producer Consumer implementation - like producing data, pushing it to Kafka and then again consuming from Kafka to send an update to the Elasticsearch datastore.
- Data Streamer:
  - Producer
  - Consumer
Initialization of AIOKafka Producer with KAFKA_INSTANCE

Producer sends message to Kafka. This message holds information related to product.
POST http://127.0.0.1:8001/producer/retail-product

```json
{
  "name": "Funny Farm House Ketchup",
  "category": "Dips and Ketchups",
  "price": 1.5,
  "stock": 3,
  "product_id": 10,
  "timestamp": "",
}
```

```
"name": "Funny Farm House Ketchup",
"message_id": "Funny Farm House Ketchup_e5c0f1b0-ac1e-44c7-92c1-1f86728a36dc",
"topic": "retail-product",
"timestamp": "2021-07-12 15:17:20.314894"
```
Initialization of AIOKafka Consumer with KAFKA_INSTANCE

Consume messages in a batch from “retail-product” topic.
```json
{
    "topic": "retail-product",
    "timestamp": "2021-07-12 20:47:21.067802",
    "product_name": "Funny Farm House Ketchup",
    "product_id": 10,
    "success": true
}
```
DATA UPDATE

Get the product doc which has to be update by “_id”

Update the product price and stock

Send update request to Elasticsearch to update the following product doc.


RESULT

```json
{
    "_index": "retail-catalogue",
    "_type": "_doc",
    "_id": "10",
    "_version": 3,
    "_seq_no": 4,
    "_primary_term": 2,
    "found": true,
    "_source": {
        "product": {
            "id": 10,
            "name": "Funny Farm House Ketchup",
            "category": "Dips and Ketchups",
            "price": 15.0,
            "stock": 5
        },
        "update_ts": 1625931472
    }
}
```

```json
{
    "_index": "retail-catalogue",
    "_type": "_doc",
    "_id": "10",
    "_version": 4,
    "_seq_no": 5,
    "_primary_term": 2,
    "found": true,
    "_source": {
        "product": {
            "id": 10,
            "name": "Funny Farm House Ketchup",
            "category": "Dips and Ketchups",
            "price": 15.0,
            "stock": 3
        },
        "update_ts": 1625931472
    }
}
```
CONCLUSION AND LEARNINGS

● We learnt about how Producer Consumer models work and their advantages to process millions of data.

● Deep dive in example of Producer Consumer working and how it can be used with as a background service.

● Understanding of how datastore is being updated and achieving high performance and other benefits using messaging queue.

● Python code implementation can be found in my git repository - https://github.com/harshit98/Retail-Updates-Streamer

● You can reach me out on Matrix / Email / Linkedin.
THAT’S ALL FOLKS, THANK YOU!

EURO PYTHON
2021  Jul 26 - Aug 1  Online

harshit98  linkedin  harshit-prasad  twitter  HarshitPrasad8  harshitprasad.com