

## پیاده‌سازی پایپ‌لاین داده با **Elasticsearch** و **Kafka**، **Logstash**

### شرح تسک

□ ارسال داده‌های خام :

□ یک فایل **JSON** یا یک لیست از دیکشنری‌ها در پایتون که شامل داده‌های خام و نامرتب است ( به عنوان مثال در آخرین صفحه ) را در نظر بگیرید. این داده‌ها ممکن است شامل فیلدهای نامعتبر یا مقادیر ناقص باشند. شما باید این داده‌ها را با یک اسکریپت **Python** به یک تاپیک **Kafka** به نام **"test\_pipeline"** ارسال کنید.

□ پردازش داده با **Logstash** :

□ داده‌های ارسال شده به **Kafka** را با **Logstash** پردازش کنید.

□ فیلدهای نامعتبر را اصلاح کرده و داده‌ها را به فرمت استاندارد (مثال در آخرین صفحه) تبدیل کنید.

□ داده‌های پردازش‌شده را در یک ایندکس **Elasticsearch** ذخیره کنید.

□ استقرار سرویس‌ها با **Docker** :

□ تمامی سرویس‌های مورد نیاز

(**Python Script**، **Kafka**، **Logstash**، **Elasticsearch**، **Kibana**) را با **Docker** و

**Docker Compose** به صورت یکجا و همزمان راه‌اندازی و اجرا کنید.

□ فایل **docker-compose.yml** را ایجاد کرده و اطمینان حاصل کنید که تمامی سرویس‌ها به

درستی با یکدیگر ارتباط دارند.

□ اعتبارسنجی داده‌ها :

□ بررسی کنید که داده‌های پردازش‌شده در **Elasticsearch** به درستی و در ایندکسی به نام

**"test\_pipeline"** ذخیره شوند.

□ از **Kibana** برای مشاهده داده‌های ذخیره شده و بررسی صحت اطلاعات استفاده کنید.

## خروجی‌های مورد انتظار:

✓ دیتای اولیه که می‌تواند به صورت یک فایل **JSON** یا یک لیست از دیکشنری‌ها در پایتون باشد.

✓ اسکریپت **Python** که داده‌های نامرتب (مورد بالا) را به **Kafka** ارسال می‌کند.

✓ پیکربندی **Logstash** شامل فیلترهای پردازش داده و خروجی به **Elasticsearch** باشد.

✓ فایل **docker-compose.yml** که برای اجرای تمامی سرویس‌ها به صورت همزمان می‌باشد.

✓ اسکرین‌شات از **Kibana** که نشان‌دهنده پردازش و ذخیره موفق دیتا در **Elasticsearch** باشد.

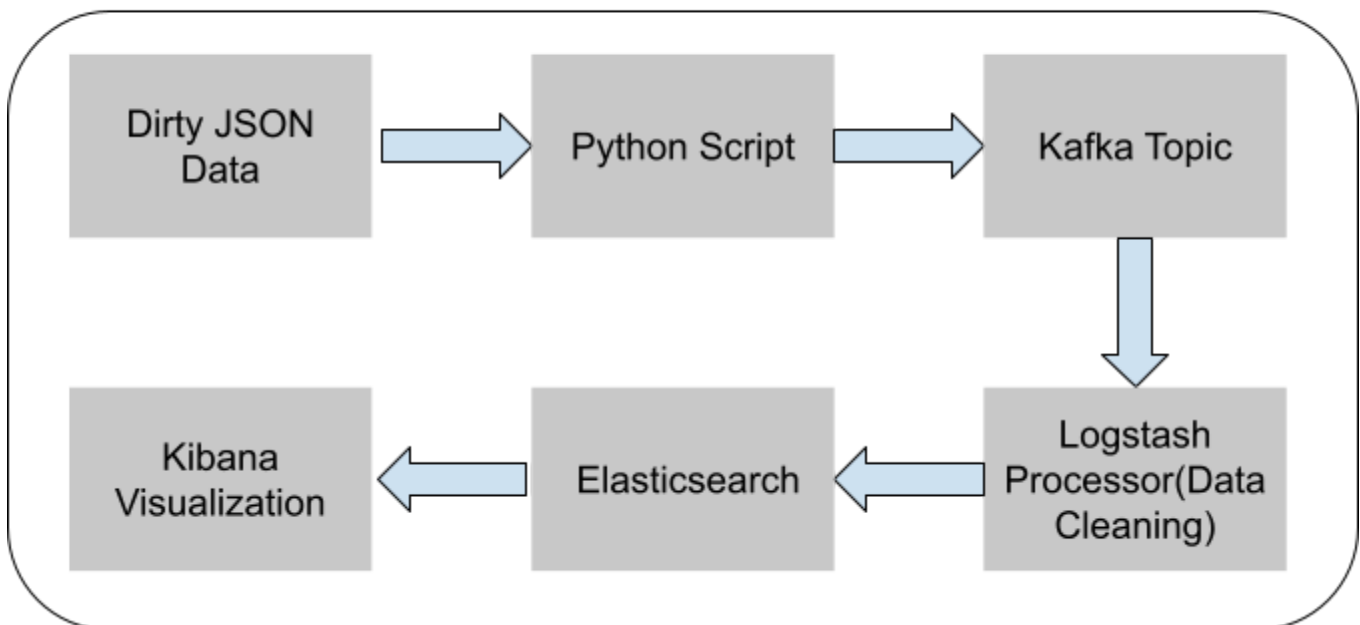
✓ مستندات کوتاه شامل توضیح مراحل پیاده‌سازی و چالش‌های مواجه شده به همراه فایل‌های نوشته شده.

✓ فایل‌های پروژه را در یک ریپازیتوری **GITHUB** ذخیره کرده و لینک ریپازیتوری را برای ما ارسال کنید.

\*\*\* توجه کنید که این فرآیند و پایپ لاین داده باید به صورت یکجا و همزمان با اجرای یکبار سرویس و

فایل **docker compose** و متعلقات آن راه اندازی شود \*\*\*

### Docker Compose Services



## مثال داده های نامرتب:

ساختار ورودی دیتای خام به اسکریپت پایتون می تواند مثل **JSON** زیر باشد

```
[  
  {"id": "123", "full_name": "Ali", "timestamp": "2025/01/31 12:34:56", "active": "true"},  
  {"id": "124", "full_name": " ", "timestamp": "2025-01-31T14:20:00Z", "active": "false",  
  "extra_field": "i should be removed!"}  
]
```

## مثال خروجی پردازش شده در **Elasticsearch**:

خروجی دیتای پردازش شده برای دیتای ورودی بالا باید مثل **JSON** زیر باشد

```
[  
  {"id": 123, "name": "Ali", "timestamp": "2025-01-31T12:34:56Z", "active": True},  
  {"id": 124, "name": null, "timestamp": "2025-01-31T14:20:00Z", "active": False}  
]
```

---

مدت زمان انجام تست: **یک هفته**

موفق باشید !