

מסמך דרישות - מיפוי מערכת רישום משתמשים:

FastAPI → Kafka → MongoDB

איפיוון שלב 2 – הזרמת JSON מקובץ (Batches)

2.1 מטרה

להוסיף ל-**API** יכולת לבצע **bulk publish** על בסיס קובץ JSON גדול הנמצא בתוך קונטינר ה-**API**, ולפרנסט את הנתונים ל-**Kafka** ב-**batches** מבוקרים.

השלב מתמקד ב:

- קריאה מקובץ JSON
- שליחה מבוקרת ב-**batches**
- ניהול לולאת פרסום תקינה
- עבודה עם JSON מקוון

2.2 קלט (Dataset)

יתווסף קובץ JSON בתוך **image** של שירות ה-**API**, לדוגמה:

`api/data/users_with_posts.json`

מבנה הקובץ:

- מערך של **Users**.

- כל User כולל מערך פנימי בשם `posts`.
 - ה-`posts` מקוונים בرمמת `theUser` בלבד (אין קיון עמוק נוסף).
- הקובץ נקרא מתוך מערכת הקבצים של הקונטינר בזמן ריצה.
-

Seed-Trigger חדש ל-Endpoint 2.3

GET /seed

תפקיד:
להפעיל תהליך שmaps את הנתונים מתוך הקובץ ל-Kafka.

דרישות חובה

1. קריאת הקובץ מתבצעת פעם אחת בלבד.

2. הנתונים מחולקים ל-`batches` של 10 משתמשים.

3. כל 5 שניות נשלח `batch` נוספת.

4. אין לשולח את כל הנתונים בבהא אחת.

מנגן לולאה

ה-`endpoint` מआחל לולאה שימושica לשולח `batches` עד לסיום הנתונים.

נדרש:

• תנאי עצירה ברור כאשר אין עוד רשותות לשילחה.

• אין להשאיר לולאה אינסוף ללא תנאי יציאה.

• ניתן למשם באמצעות `flag`, `while` פנימי, או מנגן אחר — כל עוד ההתנוגות תקינה.

לדוגמה Response

```
OK 200
}
, "status": "started"
"message": "seeding from file in batches of 10 every 5 seconds"
```

Kafka 2.4

- שימוש ב-topic הקיימ `users.registered`:

2 – Acceptance Criteria 2.5

השלב ייחשב תקין אם:

- קיימ קובץ JSON בתוך קונטינר ה-**API**.
- קריאה ל-**/seed** מפעילה זרימה מדורגת ל-**Kafka**.
- נשלחים batches של 10 כל 5 שניות.
- ה-**Consumer** מצליח לקרוא את ההודעות.
- מומלץ לוודא שכל הרשומות נקלטו בהצלחה בסיסד הנתונים.

לא ניתן מספר רשומות צפוי מראש. עליכם לבצע ולידציה עצמאית.

שלב נסס