תקציר עלילה

הצבא קיבל גישה למערכת מתקדמת לניטור והאזנה להודעות אלקטרוניות, וכל הודעה שמתקבלת נשלחת לעיבוד בזמן אמת. המטרה היא לזהות אם התוכן כולל רמזים לפעילות עוינת או מסוכנת, כגון מילות מפתח שמעידות על "בן ערובה" או "חומר נפץ". בהמשך כל הודעה חשודה נשמרת במסד נתונים ומנותחת כדי לסייע במעקב ואבטחה.

מטרת הפרויקט

הפרויקט מתמקד ביצירת יישום Flask שמקבל הודעות אימייל בפורמט JSON, מבצע עיבוד ראשוני על התוכן, מאתר מילות מפתח חשודות, ושולח ל-Kafka למטרות מעקב וניתוח של מידע חשוד.

התחלת עבודה

1. **שכפול הריפוזיטורי והגדרת סביבת עבודה**:
   `git clone https://github.com/OmerMunk/email.dispatcher`
   ודא שהפרויקט מוגדר עם סביבה וירטואלית וכל התלויות הנדרשות.
2. **הרצת ה-Flask**:
   וודא שהיישום Flask פועל על פורט 5000. המערכת שולחת בקשות ל-
   `http://localhost:5000/api/email`.

תיאור פרויקט

קבלת הודעות אימייל:
קבל הודעות בפורמט הבא:
json
Copy code

```python
{
  "email": "jeremy37@example.org",
  "username": "jonesalejandra",
  "ip_address": "215.67.111.124",
  "created_at": "2024-10-15T05:29:13.450066",
  "location": {
      "latitude": 8.5478895,
      "longitude": -135.24204,
      "city": "Port Josephburgh",
      "country": "PA"
```

```
    },
    "device_info": {
        "browser": "Mozilla/5.0",
        "os": "iOS",
        "device_id": "c4a3ce0d-4f4f-4bc9-9e94-b135e32cfe81"
    },
    "sentences": [
        "Public quickly spend hear sing.",
        "Difference nothing environmental shake decide.",
        "Natural southern what nice."
    ]
}
```

1. **שליחת הודעות ל-Kafka**:
שלח את ההודעות לנושא Kafka בשם `messages.all` ושמור עותק של ההודעות כפי שהן ב-MongoDB תחת אוסף `all_messages`.
2. **זיהוי תוכן חשוד**:
הקוד יאזין לכל הודעה שנכנסת. אם הודעה מכילה תוכן חשוד כמו "hostage" או "explos", היא תנותב לנושאי Kafka מתאימים (`messages.hostage` או `messages.explosive`).
3. **סידור תוכן מסוכן**:
בתוכן החשוד, סדר מחדש את המילים כך שהמשפט המסוכן ביותר יופיע ראשון.
4. **שמירת תוכן חשוד ב-PostgreSQL**:
שמור את התוכן החשוד ב-PostgreSQL בטבלאות המתאימות, תוך יצירת קשרים עם טבלאות אחרות במסד הנתונים:
   - `suspicious_hostage_content`: טבלה שמכילה את המשפטים החשודים הקשורים למילת המפתח "hostage".
   - `suspicious_explosive_content`: טבלה שמכילה את המשפטים החשודים הקשורים למילת המפתח "explosive".

שאלות שיש לענות עליהן

- כיצד ניתן להפיק את התוכן המלא של אובייקט לפי אימייל מסוים כולל תוכן חשוד?
- כיצד ניתן לקבל את המילה הנפוצה ביותר בהודעות החשודות לפי אימייל?

דרישות טכנולוגיות

- שימוש ב-Flask לניהול ה-API.
- Kafka להעברת הודעות בין היישומים.
- MongoDB ו-PostgreSQL לאחסון תוכן.
- חשיבות בהבנה וניהול של הודעות חשודות בזמן אמת.

פתחו נקודת קצה המביאה את כל התוכן החשוד עבור email מסוים.

ממשו זאת על ידי שימו ב join.

בונוס

פתחו נקודת קצה החושפת את מילה הכי נפוצה בכל הדאטאת השתמשו ב:

```
words_rank = Counter(['word1', 'word2']).most_common()
```

קריטריונים להערכת הפרויקט

- פונקציונאליות: 70 נקודות
- ארכיטקטורה: 10 נקודות
- סגנון קוד: 10 נקודות
- ניהול גרסאות עם Git: 10 נקודות

בהצלחה!