**תרגיל תכנותי - elasticsearch**

**הנחיות:**

**אופן עבודה:**

* **יש לעבוד ללא עזרי AI לסוגיהם.**
* **ניתן ואף רצוי להיעזר בחומר שניתן בכיתה, ועבודות קודמות שלכם בקורס הנמצאות ב-GitHub.**
* **מותר להשתמש בחיפוש בגוגל.**
* **מומלץ לנהל משימות עם טרלו (או ג'ירה למי שמעדיף).**
* **שימו לב - החלטה על מבנה התיקיות הוא חלק מהמטלה - וייבדק**
* **יש ליצור repository ציבורי מסודר ב-GitHub יש לוודא ששם הריפו תואם לדרישה (**[**שינוי שם של repository קיים**](https://drive.google.com/file/d/17p4QeiMCKDjvg5d5XT0ISzz2iJNLD8bS/view)**)**

**השם עבור המטלה הנוכחית: "Week\_12\_ElasticSearch\_Malicious\_Text"**

* **יש כמובן להתייחס לתכנון נכון, קריאות, עקרונות סוליד, תיעוד, מבנה קבצים, פונקציונליות ושימוש בגיט.**
* **יש לסיים לפתח את הקוד ולבדוק אותו לפני תהליכי קונטיינריזציה ופריסה.**
* **חובה לבנות הגדרות קונטיינר עבור כל שירות, אולם תהליך הפריסה ב openshift הינו בונוס.**
* **יש לשמור קובץ נפרד של פקודות CLI עבור כל העלאת קונטיינר לוקלי (אלסטיק, ומונולית)**
* **במידה ובצענו פריסה, יש לשמור קובץ נפרד של פקודות CLI עבור הפריסה בopenshift.**
* **חשבון Docker Hub בכדי למשוך ולהעלות image (צריך להתחבר לDockerHub דרך CLI).**
* **בחיבור לשירותי ענן ( dockerHub, openShift, MongoDB Atlas ) מומלץ להשתמש בנקודה חמה.**
* **זמן לביצוע: COMMITS לאחר השעה 0900 ביום שלישי (02/09/25) לא יחושבו בבדיקה.**
* **שיעורי תורה הם חובה.**
* **ניתן לצאת להפסקות על פי לוז שיוגדר בכיתה, אולם מומלץ להתמקד בתרגיל.**

**רקע:**

התבקשתם לבדוק האם יהיה ניתן לשלב את ElasticSearch ככלי לתחקור מהיר של הטקסטים מבסיס הנתונים האירני.

לרשותכם [קובץ נתונים](https://drive.google.com/file/d/1QvBm-SEnxAAqrVmWQBmxx-A8bx2pM_h0/view?usp=sharing) שהורד מבסיס הנתונים לצורך בניית מערכת הבדיקה.

הקובץ כולל טקסט, שדה סיווג לטקסטים אנטישמי או לא ותאריך ושעה של הכנסת ההודעה לבסיס הנתונים.

עליכם לבנות מערכת לעיבוד בסיסי המתבססת על **שירות בודד** (מונולית)

תפקידי המערכת:

1. טעינת הקובץ ל ElasticSearch (יש לבצע מיפוי מדויק של השדות, ואינדוקס שלהם)
2. זיהוי רגש בטקסט והוספתו כשדה חדש (את התהליך יש לבצע, בעזרת ElasticSearch לאחר טעינת הקבצים).
3. זיהוי מילת מפתח מתוך רשימת [כלי נשק](https://drive.google.com/file/d/1b-DU_BtRDxCEakEotI3ICdQJIRrGpG4e/view?usp=sharing) והוספת שדה הכולל את כלי הנשק שנמצאו ברשימה (את החיפוש והזיהוי יש לבצע בעזרת ElasticSearch )
4. יש **למחוק** רשומות מ ElasticSearch אשר אינן מסווגות כאנטישמיות, אין בהן כלי נשק והרגש שלהן ניטראלי או חיובי.
5. המערכת תחשוף 2 endpoints למשתמש חיצוני:
   * במידה וכל המידע עבר עיבוד - שליפת כל המסמכים המסווגים אנטישמיים ויש להם כלי נשק כלשהו, אחרת יוחזר ג'ייסון עם הודעה מתאימה
   * במידה וכל המידע עבר עיבוד - שליפת כל המסמכים להם 2 כלי נשק או יותר, אחרת יוחזר ג'ייסון עם הודעה מתאימה.