

## תרגיל בית – הדסים 5

### חלק א'

**סעיף א'-הגדרה:**  $M$  מספר השורות ו- $E$  מספר השגיאות השונות בקובץ הקלט (סעיף 5). ניתוח סיבוכיות זמן:

1. פיצול הקובץ קלט- עוברים על השורות בקובץ ושומרים אותם בקבצים נפרדים לכן יוצא  $O(M)$ .
2. סופרים את השגיאות לכל קובץ קטן. קובץ קטן בגודל  $100,000 \approx O(1)$ , עוברים על כל הקבצים הקטנים שיש בערך  $O(M) \approx O(M/100,000)$  ולכן יוצא  $O(M)$ .
3. מיזוג התשובות יקח לנו כאורך קובץ הקלט כפול השימוש בפונקציה  $update$  של  $Counter$  שהיא בממוצע לוקחת  $O(1)$  ולכן נקבל  $O(M)$ .
4. מציאת ה- $N$  השכיחים- החיפוש ל- $N$  השכיחים יקח  $O(E \log N)$ .
5. סה"כ יקח לנו  $O(M) + O(E \log N) \approx O(M) + O(M) + O(M) + O(E \log N)$ .
6. אם  $M \gg E$  אז החלק של מציאת ה- $N$  השכיחים יקטן ממש.
7. אם  $M = E$  אז נקבל  $O(M \log N)$  (המקרה הגרוע).

ניתוח סיבוכיות מקום:

1. מערך יחיד (השימוש חוזר) והקבצים הקטנים תופסים סה"כ  $O(M)$  מקום.
2. מבנה נתונים- לכל קובץ קטן היה  $counter$  בגודל לכל היותר  $O(100,000)$  כי זה הגודל של הקובץ הקטן ולכן פה זה  $O(1)$  זיכרון.
3. וה- $counter$  הראשי הגדול יהיה בגודל כמספר השגיאות שיש ולכן יוצא  $O(E)$ .  
סה"כ המקרה הגרוע ש  $M = E$  אז יהיה  $O(M)$  מקום.
- אבל במקרה הלא גרוע ש  $M \gg E$  אם לא מחשיבים את הקבצי הטקסט שיצרנו יהיה לנו  $O(U)$  מקום.

### סעיף ב'

3. נקבל מבנה נתונים (לדוג' מילון) שיתעדכן בלייב ובו תישמר התשובה הסופית.  
על ה-  $key$  יהיה ה- $timestamp$  (תאריך והשעה) מעוגל (לפי שעה) ובנוסף שני ערכים, את הממוצע (mean) של אותה שעה וכמה  $timestamp$ ים (count) הגיעו של אותו תאריך ושעה וכך שתגיע רשומה חדשה, לפי ה- $timestamp$  שלה נוכל לדעת איפה היא במילון נוסיף אחד ל- $count$  ונחשב את הממוצע מחדש כך: (הממוצע הישן\*(count-1) + הערך החדש)/count.
4. יתרונות: מבנה האחסון העמודי. בניגוד לפורמט csv שהוא מבוססי שורות. הוא מסדר את הנתונים לפי עמודות ולא שורות.  
זה גורם ל: שאילתה יעילה: שאילתות המתמקדות בעמודות ספציפיות יכולות לאחזר נתונים מהר יותר מכיוון שרק העמודות הנדרשות נקראות.  
דחיסה טובה יותר: נתונים דומים המאוחסנים בעמודות דוחסים בצורה יעילה יותר מאשר נתונים מעורבים בשורות.

עלויות אחסון מופחתות: גדלי קבצים קטנים משמעותית בהשוואה לפורמטים מבוססי שורות.

בנוסף היתרונות שלו, אימות נתונים: מבטיח שהנתונים תואמים לסוגים מוגדרים מראש, ומפחית שגיאות עיבוד.

יעילות משאבים: מפחיתה את העומס על המעבד והזיכרון במהלך עיבוד הנתונים.

## חלק ג'

1. 3

2. יש כמה שיטות לשידור בין שלט למזגן, 2 השיטות הנפוצות הן אינפרא-אדום ו-וויפי/בלוטות'. כמובן שלכל שיטה יש את היתרונות והחסרונות שלה, למשל לאינפרא-אדום היתרון הבולט הוא מאוד זול ופשוט אבל טוב לטווחים קצרים ומצריך מהשלט להיות בקו ראייה ישיר עם המזגן, מאידך בחיבור וויפי/בלוטות' השליטה מתבצעת מרחוק ואף ניתן לחבר כמה מזגנים, החיסרון שצריך חיבור טוב לוויפי/בלוטות אחרת לא יעבוד.
- השלט זקוק למקור מתח חשמלי (בדר"כ סוללות), מעגל חשמלי שנסגר באופן שונה עבור כל כפתור וכפתור, סוג של מעבד שיבין איזה כפתור נלחץ ומה לשלוח למזגן והאם מדובר באינפרא-אדום.
- המזגן מצוייד בקולט ( אינפרא-אדום או אמצעי לקליטת וויפי/בלוטות), מעבד שמבין איזו פקודה נשלחה לו ולבצע אותה. המזגן מכיל חלקים מכניים כמו מנוע, מדחס, מאוורר ותרמוסטט שאת פעילותם הוא משנה לפי הפקודה שהתקבלה.
- האינפרא-אדום שיוצא מהשלט נשלח באופן של קידוד ארוך ומהיר של אותות אור/חוסר אור למזגן. המזגן קולט אותות אלה ויודע לתרגם את בקשת המשתמש לשינוי בפעילות המזגן.
- בשיטת השידור של וויפי/בלוטות' נשלחת פקודה טקסטואלית ובה המזגן יודע לקבל את הפקודה, להבין אותה (הוא יותר סוג של מחשב) וככה הוא יכול לבצע שינויים בפעילות המזגן.