**תרגיל בית – הדסים 5**

**חלק א'**

**סעיף א'**

5. ניתוח סיבוכיות זמן:

1. פיצול הקובץ קלט- נגדיר שמספר השורות הוא M, עוברים על השורות ושומרים אותם בקבצים נפרדים לכן יוצא .
2. סופרים את השגיאות לכל קובץ קטן. קובץ קטן בגודל ועוברים על כל הקבצים הקטנים שיש בערך ולכן יוצא .
3. מיזוג התשובות יקח לנו כאורך קובץ הקלט כפול בשימוש בפונקציה update של Counter שהיא בממוצע לוקחת ולכן יצא לנו כאן .
4. מציאת ה-N הנפוצים- נגדיר E הוא מספר השגיאות השונות שיש לנו בקובץ, כלומר החיפוש לN הכי נפוצים יקח ).

סה"כ יקח לנו

*אם אז החלק של מציאת ה-N הנפוצים יקטן משמעותית*

*אם אז נקבל (המקרה הגרוע)*

ניתוח סיבוכיות מקום:

1. *הקבצים הקטנים תופסים סה"כ מקום.*
2. *מבנה נתונים- לכל קובץ קטן היה counter בגודל לכל היותר כי זה הגודל של הקובץ הקטן ולכן פה זה זיכרון.*
3. *והcounter הראשי הגדול יהיה בגודל כמספר השגיאות שיש ולכן יוצא .*

*סה"כ המקרה הגרוע ש אז יהיה מקום.*

*אבל במקרה הלא גרוע ש ואם לא מחשיבים את הקבצי הטקסט שיצרנו יהי לנו מקום.*

**סעיף ב**'

3. יהיה לנו מבנה נתונים (לדוג' מילון) שבו נשמור את התשובה הסופית.

הוא ישמור לנו את הtimestamp (תאריך והשעה) המעוגל לפי שעה ובנוסף שני ערכים את הממוצע(mean) של אותה שעה וכמה timestampים (count) הגיעו של אותו תאריך ושעה וכך שמגיע רשומה חדשה לפי הtimestamp שלה נוכל לדעת איפה היא במילון נוסיף אחד לcount ונחשב את הממוצע מחדש כך: (הממוצע הישן\*(count-1) + הערך החדש)\count.

4. יתרונות: מאפיין ההיכר שלו הוא מבנה האחסון העמודי. בניגוד לפורמטים מבוססי שורות כמו csv. 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.

תיאור עצמי: כלים ומסגרות יכולים לפרש את הקובץ בקלות מבלי לדרוש הגדרות סכימה חיצוניות.

שיפור ביצועים: ממזער פעולות קלט/פלט, ומאיץ שאילתות.

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

**חלק ג'**

1. 2
2. a. מה​ו אופן השידור ‌בין השלט למ​זגן?

b.

c.