

ממך 12

קורס מימוש מערכות בסיסי נתונים, 20574

מגיש טל גלנצמן, 302800354

תאריך 05/04/2021 סמסטר 2021ב

שאלה 1

סעיף א

בדף יש 10 רשומות של האינדקס E1 כלומר, כאשר סורקים את האינדקס נידרשת גישה אחת לזכרון עבור איתור כל 10 רשומות ביחס P. לכן, על מנת לאתר את הרשומה עבורה id=567 נצטרך

$$\lceil \frac{567}{10} \rceil = 57$$

גישות לדיסק. רק נציין, שקריאת הרשומה עצמה דורשת כמובן גישה נוספת.

סה"כ, נדרשות 57 גישות לדיסק על מנת לאתר את הרשומה המבוקשת

סעיף ב

רשומה של האינדקס E2 מצביעה על רשומה ראשונה של האינדקס E1 בדף ובכך למעשה מצביעה ל-10 רשומות רצופות של E1. על מנת לאתר את הרשומה המבוקשת ביחס P, נצטרך לגשת לדף ה-57 באינדקס E1 ולאתור דף זה נצטרך לגשת לדף החמישי באינדקס E2 עפ"י החישוב

$$\lceil \frac{57}{10} \rceil = 6$$

משמע, נדרשות 7=6+1 גישות לדיסק על מנת לאתר את הרשומה המבוקשת

סעיף ג

עבור השדה salary יש 50 ערכים שונים המתפלגים אחיד - נסיק אם כן כי עבור כל ערך $s \in salary$ קיימות בממוצע $\frac{1000}{50} = 20$ רשומות עבורן השדה salary מקבל את הערך s. נסיק אם כן כי 2 הדפים האחרונים של האינדקס E3 מכילים רשומות הצבעה לרשומות של P אותן אנו מחפשים.

אם נבצע סריקה על האינדקס מהערך הקטן לערך הגדול נצטרך למעשה לסרוק את כל האינדקס מהשידורש 100 גישות לדיסק - עבור קריאה של 100 דפים.

גם, באופן מעשי, אומנם התפלגות הערכים אחידה, אבל עדיין קיימת הסתברות לא אפסית שבה כל הרשומות מקבלות אותו ערך, לכן גם כאן, במקרה הגרוע, נצטרך לגשת לכל דפי האינדקס מהשידורש 100 גישות לדיסק.

הערה אם ננצל את העובדה שהאינדקס ממורן, ואנו מחפשים את הערך המקסימלי של salary נובצע סריקה מהסוף להתחלה נצטרך לבצע $\frac{20}{10} = 2$ גישות לדיסק בממוצע

סעיף ד

הרמה הראשונה של האינדקס E4 תכלול רשומת מצביע עבור כל ערך אפשרי של salary, כלומר 50 רשומות הצבעה מה שיידרוש $\frac{50}{10} = 5$ דפים. המצביע לרמה השנייה של האינדקס ימצא בדף החמישי ולכן כדי להגיע למצביע זה יידרשו 5 גישות לדיסק.

כפי שראינו בסעיף ג', עבור כל ערך של salary יהיו בממוצע 20 שורות המקבלות ערך זה, ולכן, ברמה השנייה של האינדקס יידרשו שני דפים להכיל את רשומות המצביעים לרשומות אלו.

בפרט זה נכון עבור הערך המקסימלי של salary, לכן יידרשו 2 גישות לדיסק על מנת למצוא את הרשומות הדרושות כאשר נתון לנו כבר גפי הרמה השנייה.

לסיכום נדרשות 7 גישות לדיסק - 5 גישות לדיסק על מנת למצוא את הדף הראשון בשרשרת הדפים, ועוד שתי גישות לדיסק על מנת לקרוא את שני הדפים הרלוונטיים המכילים את המצביעים לרשומות עצמן

שאלה 2

סעיף א

$$N = 10^8, S = 4, P = 4096$$

נפח האחסון של עמודה A_1 הוא NS בתים, כלומר

$$\lceil \frac{NS}{P} \rceil = 97657$$

דפים.

עבור כל אחת מהעמודות $A_{2,3,4,5}$ נדרשים $2NS$ בתים שכן זו מחזיקה גם את ערך העמודה וגם את ערך המפתח, כלומר

$$\lceil \frac{2NS}{P} \rceil = 195313$$

דפים.

כל עלה בעץ האינדקס מחזיק ערך של A_1 ועוד 4 מצביעים, כלומר 20 בתים.

סה"כ

• אם לא נרצה לאחסן את העמודה A_1 כשלעצמה נפח האחסון יהיה $4 \cdot 195313 = 781252$ דפים

• אם כן נרצה לאחסן את העמודה A_1 כשלעצמה נפח האחסון יהיה $781252 + 97657 = 878909$ דפים

על זה נוסיף את נפח אחסון האינדקס - לא בדיוק הבנתי איך ניתן לחשב

סעיף ב

נפרק את השאלתא לשלבי ביצוע

1. מציאת הערך המינימלי של A_4
2. מציאת הערך המינימלי של A_2 אשר גדול ממש מהערך המינימלי של A_4
3. סריקת A_2 החל מהערך שהתקבל בשלב הקודם

תחילה נציין שנפח האיחסון עבור כל עמודה הוא ידוע ולכן ניתן לגשת באופן ישיר לכל אחת מהרשומות - הכוונה לפי סדר, לא לפי ערך.

נסמן $K = 195313$ מספר הדפים לאחסון עמודות ערכים

שלב 1 מציאת הערך המינימלי של A_4 מתבצעת פשוט ע"י קריאת הדף הראשון של A_4 ושליפת הערך הראשון, כיוון שהעמודה ממוינת, זהו הערך המינימלי. תידרש כאן גישה אחת לדיסק.

שלב 2 כעת נבצע חיפוש בינארי על A_2 למציאת הערך המינימלי של A_2 המינימלי אשר גדול מהערך המינימלי של A_4 אשר קיבלנו בשלב הקודם. בממוצע תהליך זה ייקח $\lceil \log_2 K \rceil = 18$ גישות לדיסק

שלב 3 בהנחת בתפלגות אחידה, נניח שהערך של A_2 שהתקבל בשלב הקודם הוא בחצי העמודה. בממוצע יידרשו $\lceil \frac{K}{2} \rceil = 97657$ גישות לדיסק - אבל, בשלב הקודם, לפי שיטת החיפוש, כחצי הגישות בוצעו על דפים שבהם הערך גדול מהערך שחופש ולכן 9 דפים הוטמנו בדפי החוצץ. כאן אנחנו ניגשים לפחות מ-100000 דפים לכן סביר להניח ש9 הדפים האלו עדיין מוטמנים. נסיק כי יידרשו כאן כ- 97648 גישות לזכרון

בסה"כ בממוצע יידרשו כ- $97648 + 18 + 1 = 97667$ גישות לדיסק

שאלה 3

שאלה 4