

אינדקסים ופעולות צירוף

הנחות כלליות

- שמירת בלוקים בזכרון הראשי (RAM) מוגבלת ל-2 בלוקים בבת אחת.
- לא ניתן לבצע פעולת צירוף של יותר משתי רלציות בו זמנית. יש לשמור בדיסק את תוצאת הצירוף בין שתי רלציות כתוצאת ביניים, לאחר מכן להתייחס אליו כרלציה חדשה ולבצע פעולות צירוף נוספות.

סימונים

- $R \bowtie S$ - פעולת הצירוף שרוצים לבצע.
- B_R - מספר הבלוקים הדרושים לאחסון רשומות R .
- N_R - מספר רשומות R .

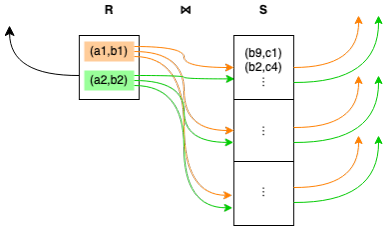
ללא אינדקסים (Nested-Loop Join)

המימוש הפחות נאיבי

עבור כל בלוק b_R ב- R :
 עבור כל בלוק b_S ב- S :
 עבור כל רשומה t ב- b_R :
 לכל רשומה s ב- b_S , פלוט (t, s)
 אם t, s מקיימים את תנאי הצירוף.
 סיבוכיות: $O(B_R \cdot B_S)$

המימוש הנאיבי

עבור כל בלוק b_R ב- R :
 עבור כל רשומה t ב- b_R :
 עבור כל בלוק b_S ב- S :
 לכל רשומה s ב- b_S , פלוט (t, s)
 אם t, s מקיימים את תנאי הצירוף.
 סיבוכיות: $O(B_R + N_R \cdot B_S)$



אינדקס עץ $B+$ (Indexed Nested-Loop Join)

הנחות בסיס

- כל גישה לצומת בעץ דורשת טעינה של בלוק מהדיסק.
- בעץ $B+$ מסדר d , כל צומת מלבד השורש מחזיקה $d \leq k \leq 2d$ מפתחות (בשורש $1 \leq k \leq 2d$).
- עלי העץ הם בלוקים, בכל בלוק $d \leq k \leq 2d$ מפתחות עם הפנייה למיקום הבלוק שמכיל את הרשומה בדיסק. כל עלי העץ מרכיבים רשימה מקושרת דו-כיוונית.

* בפעולות אלה יתכן off by one בהתאם להזחה של התוצאות בבילוקי העלים (מסומן באדום)

פעולת צירוף ע"י אינדקס

עבור כל בלוק b_R ב- R :

עבור כל רשומה t ב- b_R :

אתר ב- S ע"י האינדקס את כל הרשומות s

כך ש- t, s מקיימים את תנאי הצירוף ופלוט (t, s) .

□ סיבוכיות בפועל: $O(B_R + |R \bowtie S|)$

□ חישוב I/O :

סריקת טווח ברלציה R ע"י אינדקס

- חפש את בלוק העלה שמתייחס לחסם העליון/התחתון על הטווח בעץ.

- כל עוד לא זוהתה חריגה מהטווח ועדיין יש עלים רלוונטיים:

* פלוט את המפתחות החוקיים בעלה.

* אם עדיין לא זוהתה חריגה, טען את בלוק העלה הבא בכיוון המתאים.

□ חישוב I/O במקרה הגרוע:

$$\lceil \log_d(N_R) \rceil + \left\lceil \frac{(\text{מספר רשומות})}{d} \right\rceil + 1?$$

- האלגוריתם רק פולט את המפתחות של רשומות בטווח ולא שומר אותם בדיסק.

- כדי לקבל שדות מ- R שלא נמצאים במפתח, יש לטעון כל בלוק מתאים מהדיסק.

- כדי להשתמש במידע לצירופים נוספים, יש לבצע כתיבה של תוצאת ביניים.

$$B_R + \sum_{t \in B_R} \left(\lceil \log_d(N_S) \rceil + \left\lceil \frac{(\text{מספר רשומות } t \text{ תוצאה עבור } d)}{d} \right\rceil + 1? \right)$$

אם מספר בלוקי תוצאה קבוע C לכל רשומה,

החישוב יהיה $B_R + N_R \cdot \lceil \log_d(N_S) \rceil \cdot C$

- אלגוריתם זה רק פולט את המפתחות ב- S לפי האינדקס ולא שומר אותם בדיסק.

- כדי לקבל שדות מ- S שלא נמצאים במפתח,

יש לטעון כל בלוק מתאים מהדיסק.

- כדי להשתמש במידע לצירופים נוספים, יש לבצע כתיבה של תוצאת ביניים.