

אוניברסיטת בן-גוריון

המחלקה להנדסת מערכות תוכנה ומידע

קורס: בסיסי נתונים – תשפ"ג

מטלה 4 – אופטימיזציה

מתרגל אחראי: אדם עוזיאל.

שאלה 1:

נתונות טבלה R לא ממוינת שגודלה 100 בלוקים וטבלה S שגודלה 1000 בלוקים. מעוניינים לבצע Hash-Join בין שתי הטבלאות. לא מוגדרים אינדקסים על הטבלאות. הזיכרון הפנימי כולל מקום פנוי ל-5 בלוקים.

1. (1 נק') לכמה מקסימום קבוצות יכולה פונקציית הגיבוב לחלק את הטבלאות בשלב א של HashJoin בהינתן מגבלת הזיכרון הנתונה?
2. (2 נק') ידוע כי החלוקה בשלב א' היא מאוזנת לחלוטין. כמה עוד מקום פנוי בזיכרון נדרש במדויק אם מעוניינים להשלים את שלב א ואת שלב ב.
3. (7 נק') בהנחה כי לא ניתן להקצות את הבלוקים הנוספים – הציעו שיטה חלופית יעילה לביצוע שלב ב' (הניחו ששלב א של HashJoin בוצע בהצלחה במגבלת הזיכרון). מה הסיבוכיות במונחי פעולות קריאה וכתיבה הנדרשות להשלמת הפעולה?

שאלה 2:

נתון בסיס נתונים לניהול הרישום של סטודנטים לקורסים שונים הכולל את הטבלאות הבאות:

Students(std_id, first_name, last_name, city)

Courses(course_id, course_name, points)

Enrolment(course_id, std_id, year, semester)

וכן השאילתא הבאה:

SELECT last_name

FROM Students NATURAL JOIN Enrolment NATURAL JOIN Courses

WHERE course_name = "Databases" AND city = "Beer-Sheva"

א. שרטטו עץ שאילתא ראשוני (2 נק')

ב. שרטטו עץ שאילתא אופטימאלי ככל האפשר בהתבסס על כללי אופטימיזציה היוריסטית. (8 נק').

ג. הסבירו מה צריך לשנות בפעולת Hash – Join כדי לאפשר ביצוע הפעולה (3 נק').

S LEFT OUTER JOIN R

ד. בהמשך לסעיף ג' הסבירו כיצד, אם בכלל, משתנה התנאי לכמות הזיכרון המינימאלית הנדרשת לביצוע הפעולה (2 נק')

הוראות הגשה:

1. יש להציג דרך פתרון מלאה לכל אחת מהשאלות.
2. יוגש קובץ PDF בצורה הבאה: ID1_ID2.
3. ההגשה היא בזוגות בלבד, אלא אם כן התקבל אישור מיוחד מסגל הקורס. כמו כן, רק אחד מבני הזוג אמורה להגיש את העבודה.
4. תאריך ההגשה: 12.1.2022.

בהצלחה!