שאלה 1

|  |  |  |
| --- | --- | --- |
|  | Employees(eid, ename, did, rank, sal) | Departments(did, mgrid, budget) |
| גודל רשומה | 40 byte | 20 byte |
| התפלגות ערכים | sal: [1…1,000,000] | budget: [1…1,000,000] |
| מס' רשומות | 100,000 | 5,000  2500 ניירובי, 2500 מנילה |
| עידכון? |  | לעיתים רחוקות |
| שאילתות | קצב זהה | קצב זהה |
|  | 1. 50% - שם מנהל מחלקה באתר (JOIN) 2. 50% מניירובי – יחס ממוצע בין משכורת מנהל מחלקה ותקציב המחלקה (70% מהן – נתונים מניירובי, 30% נתונים ממנילה) (JOIN) 3. 50% ממנילה – שם מנהל, דרגה ממוצעת במחלקה (JOIN) – לכל מחלקה הסת' שווה להופיע כארגומנט (50% לכל אחת) | |

Page I/O cost: td, Page delivery cost: ts

join: sort-merge join

1. שיטה לביזור נתונים – שכפול נתונים מותר.

חלוקה אופקית (לפי תנאים): שלמות, שחזור, הפרדה – minterm predicate

כיוון שכל השאילתות מכילות JOINים, **שיטת Derived HF – עם SEMIJOIN** – עדיפה.

חלוקה אנכית (לפי שדות): יש צורך לבצע חישובים שאין להם מספיק נתונים, אך ננתח לפי שליפה.

מכאן:

נבחין כי אין נתון המכיל את שם האתר, לכן לא נוכל להשתמש בפרדיקט שמכיל את שם האתר.

1. עבור מחצית מהשליפות (שליפה מסוג מס' 1 בטבלה) נשמור את הנתונים **בשכפול מלא** של הרשומות (מבחינת חלוקה אופקית).

מבחינת חלוקה אנכית – נרצה לשמור רק את האטריבוטים:

Departments(did, mgrid(, Employees(eid, ename)

**נשים את תתי הרלציות האלו בשני האתרים**

1. עבור ניירובי (שליפה 2) : חלוקה אופקית: אין חלוקה של פרדיקטים, לכן נשים את כל הרשומות

חלוקה אנכית: נצטרך את האטריבוטים הנוספים:

Employees(sal), Departments(budget)

וכך נרחיב את הרלציות מבחינה אנכית רק עבור ניירובי.

1. עבור מנילה (שליפה 3) : חלוקה אופקית: אין חלוקה של פרדיקטים, לכן נשים את כל הרשומות

חלוקה אנכית: נשים לב שאנחנו מוסיפים לשליפה 1 שדה נוסף – דרגה ממוצעת. לכן נוסיף את האטריבוטים הבאים:

Employees(did, rank)

כלומר, בניירובי נשים:

Departments(did, mgrid,budget(, Employees(eid, ename,sal)

במנילה נשים:

Departments(did, mgrid(, Employees(eid, ename, did, rank)

רלציית המחלקות מתעדכנת לעיתים רחוקות – לכן ניתן לפצל אותה אופקית לשני האתרים ללא חשש

1. כעת לא ניתן לשכפל נתונים, מכאן שהתשובה תשתנה – חפיפה אפשרית באטריבוטים תהיה רק המפתח. נכון לעכשיו כל הרשומות נמצאות בשני האתרים ואין פרדיקטים שמאפשרים לפצל לפי רשומות.

לכן נותר רק לפצל לפי אטריבוטים. בטבלת Departments(did, mgrid) יש חפיפה של mgrid ובטבלת Employees(eid, ename) יש חפיפה של ename. שני שדות אלו נמצאים בחצי מכל השאילתות ובשאר השאילתות רק עבור מנילה. לכן נשים אותם רק במנילה – מכאן שעבור חצי מהשאילתות מניירובי נצטרך להמתין עד שהנתונים יעברו באינטרנט.

כלומר, בניירובי נשים:

Departments(did,budget(, Employees(eid,sal)

במנילה נשים:

Departments(did, mgrid(, Employees(eid, ename, did, rank)