

TDT4145 Datamodellering og databasesystemer

Queryprosessering

- a) **Flettesortering:** Anta du har ei heapfil på 1 TB som du skal sortere med 1 GB tilgjengelig buffer. Heapfila er organisert i blokker på 32 KB. Hvor mange blokker må leses og skrives for å sortere fila?
- b) **Aksessveier:** Anta du har en tabell `Ansatt` lagret som en heapfil med 2000 blokker med 100 poster i hver blokk. Regn ut hvor mange I/Oer du får ved det følgende queriet:
`SELECT * FROM Ansatt WHERE ansattId = 32167;`
- c) Regn ut det samme hvis den samme tabellen er lagret i et Clustered B+-tre. Søkenøkkelen i B+-treet er `ansattId`. Anta blokkene er 8 KB store og at indekspostene (postene på nivå 1 og høyere) har 10% av størrelsen av en `Ansatt`-post på løv nivå.
- d) Gjør den samme beregningen som i b), men denne gang med queriet:
`SELECT * FROM Ansatt WHERE etternavn < 'Hansen' AND etternavn > 'Davidsen' ;`
Anta 10% av postene kvalifiserer i dette queriet.
- e) Gjør omtrent som i beregningen som i c), men denne gang med queriet fra d). Denne gangen bruk et B+-tre med `etternavn` som nøkkel.
- f) Anta tabellen er lagret som ei heapfil som i b), men det finnes en indeks, et (unclustered) B+-tre med søkenøkkel `'etternavn'`. Anta indeksposten med `(etternavn, RecordId)` trenger 20% av plassen som en `ansattpost` trenger. Regn ut antall I/O for å evaluere queriet i d).
- g) **Join:** Vi har følgende to tabeller
`Employee(ssn, first_name, last_name, bdate, sex, salary, super_ssn, dno)`
`Department(dname, dnumber, mgr_ssn, mgr_start_date)`
Anta `Department` har 500 poster lagret i 10 diskblokker og `Employee` har 60000 poster i 2000 diskblokker.
Gitt det følgende queryet:
`SELECT e.last_name, e.first_name
FROM Department d, Employee e
WHERE e.dno=d.dnumber AND
d.dname='Accounting' ;`

Anta du har 12 diskblokker tilgjengelig i buffer. Hvor mange I/Oer får du ved å bruke en nested-loop-join her for joinen `e.dno=d.number`?
- h) Er det noe annet i dette queriet som kan hjelpe oss å redusere mengden poster som kan joines?