

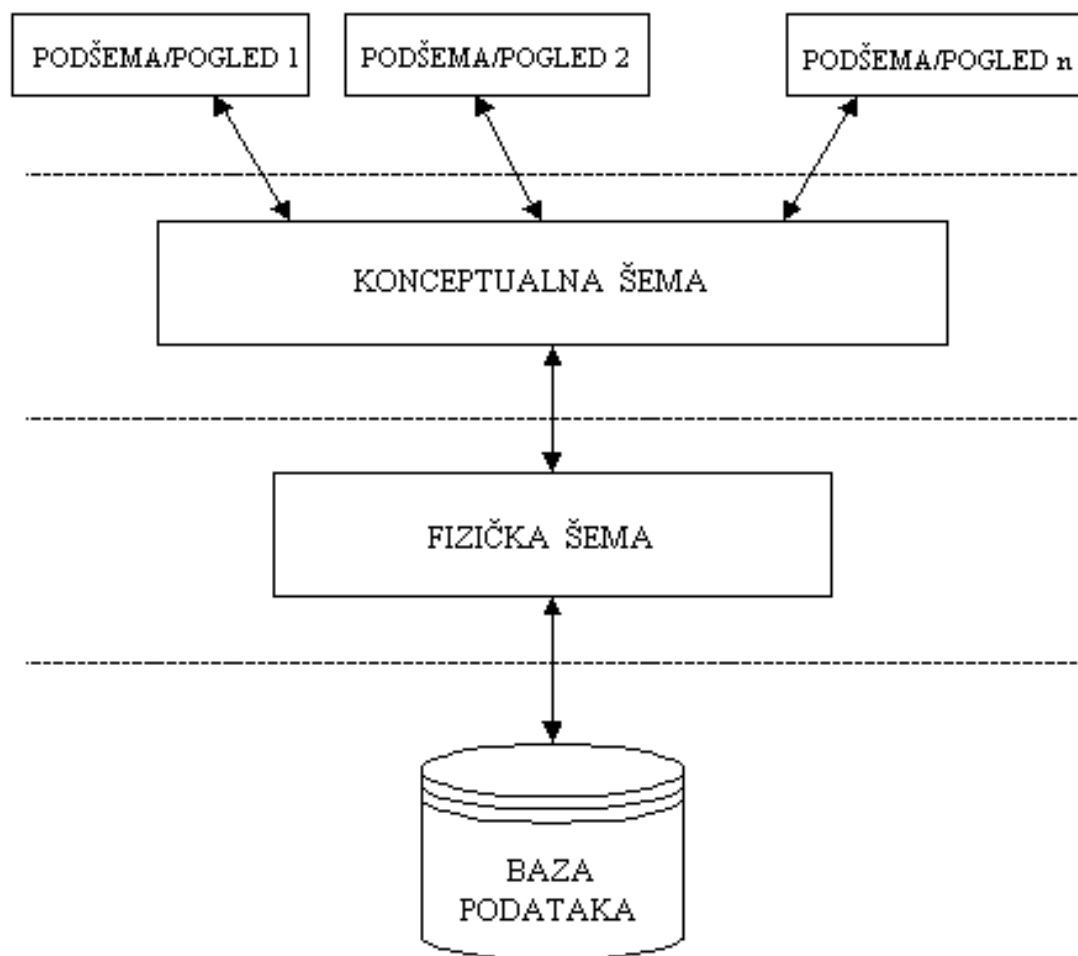
Administracija baza podataka (3)

Fizičko projektovanje baze podataka

Faze projektovanja BP

1. Prikupljanje i analiza zahteva
 2. Logičko projektovanje baze podataka
 3. Izbor sistema za upravljanje bazom podataka
 4. Prevođenje modela podataka
 5. **Fizičko projektovanje BP**
 6. Implementacija BP
-

Arhitektura tronivojskog SUBP

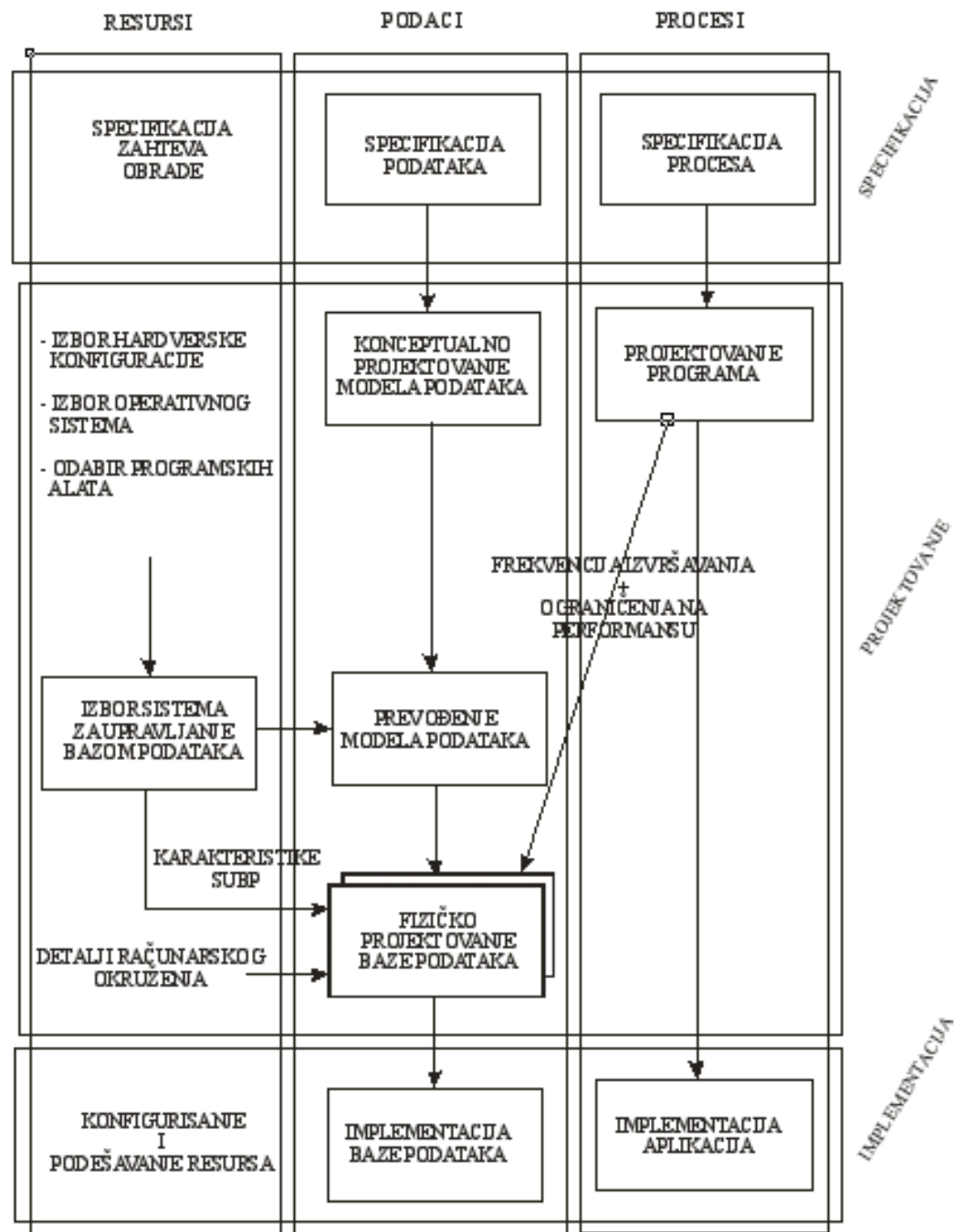


Osnovna karakteristika savremenih SUBP:

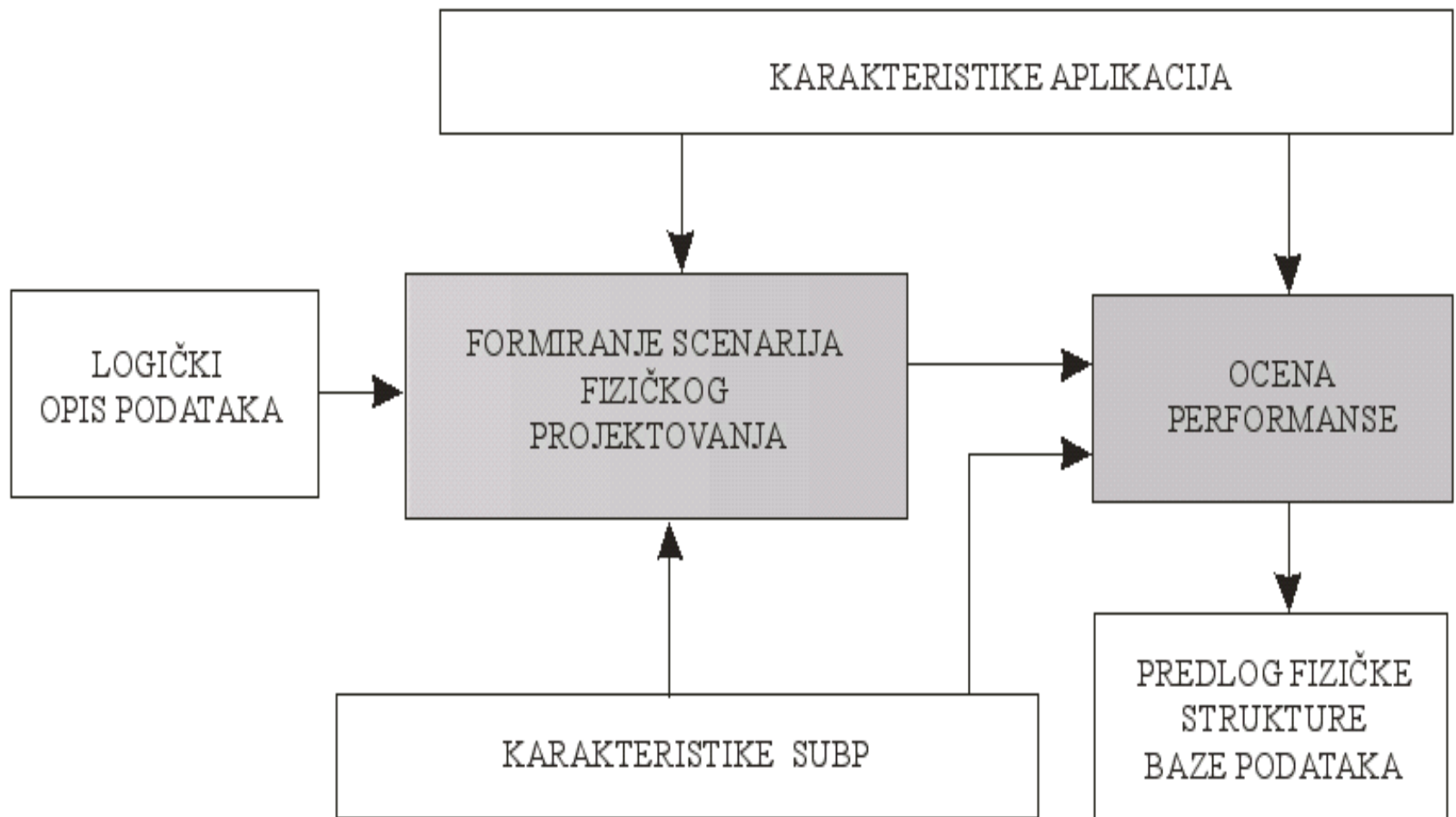
- Fizička nezavisnost podataka - razdvajanje logičkih tipova podataka i njima pridruženih operacija od fizičke reprezentacije

Posledica:

- Logički model podataka može se predstaviti sa više različitih fizičkih struktura podataka
-



Blok šema fizičkog projektovanja

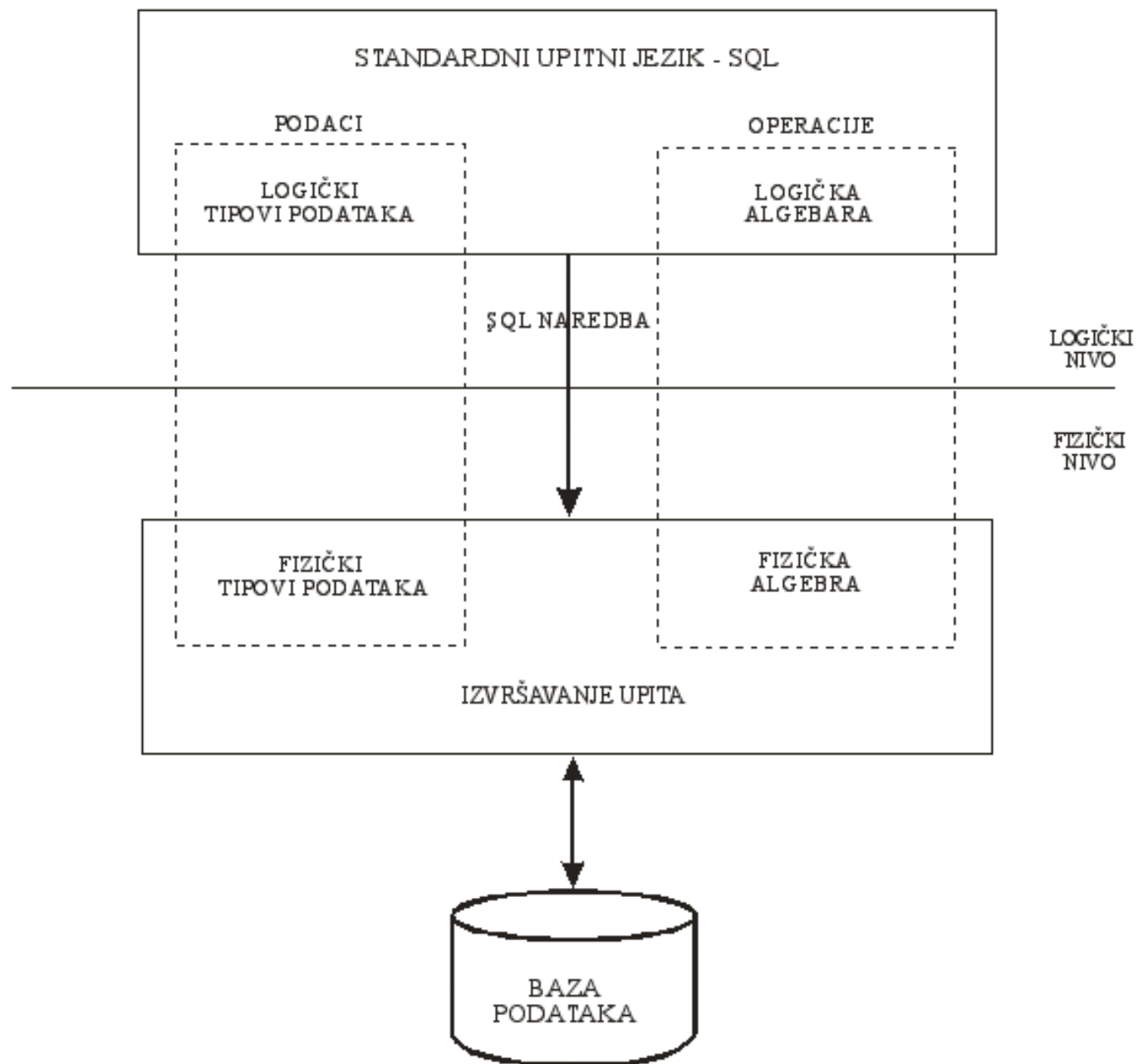


Karakteristike relacionih SUBP

- ❑ Neproceduralni upitni jezik**
 - ❑ SUBP obezbeđuje mehanizam da se dođe do traženih podataka**
 - ❑ Algoritmi obrade**
 - implementacija relacionih operacija
 - implementacija funkcija (agregacija, grupisanja, eliminacija duplikata ...)
-

Karakteristike relacionih SUBP

- ❑ **Zadatak sistema je izbor optimalnog načina izvršavanja upita**
 - Za dati skup algoritama obrade i
 - Projektovanu fizičku strukturu baze podataka
-



Karakteristike relacionih SUBP bitne za fizičko projektovanje

- ☐ Fizičke strukture podataka.
 - ☐ Implementacija osnovnih relacionih operacija (Selekcije, Projekcije, Operacija nad skupovima)
 - ☐ Implementacija operacije Spoja
 - ☐ Način optimizacije upita
 - ☐ Ažuriranje (*n*-torki i *indeksa*)
-

Fizičke strukture podataka

- ❑ Kakvu fizičku organizaciju podataka podržava konkretni (odabrani) SUBP.
 - Kako je fizički organizovan Prostor baze podataka kod konkretnog (odabranog) SUBP.
 - Kakav koncept grupisanja podataka podržava (***File Group***-e, ***Tablespace***-eve, *Data set*-ove....).
 - Koje vrste pristupnih struktura za brže pristup (pronalaženje) podataka u bazi podržava.
 - Koje vrste indeksa konkretni SUBP podržava i koju fizičke strukture podataka konkretni SUBP koristi.
 - Da li podržava heširanje - pozicija (fizička) podatka se određuje na osnovu njegove vrednosti
-

Implementacija operacije SELEKCIJE

- ❑ **Za jedan uslov selekcije** koristi se
 - pristupna putanja, ili
 - linearno pretraživanje
 - ❑ **za složeni konjuktivan uslov** (log. i) izbor putanje se vrši na osnovu
 - *Faktora selektivnosti* pojedinačnih uslova
 - ❑ **za složeni disjunktivan uslov** (log. ili) koristi se
 - linearno pretraživanje, ili
 - Pristup preko više putanja i formira se unija *IDN*-ova
-

Implementacija operacije PROJEKCIJE

- ☐ Jednostavna kad *<lista obeležja>* sadrži primarni ključ;
 - ☐ u suprotnom mora se vršiti eliminacija duplikata; vrlo često se koristi heširanje
-

Implementacija operacija nad skupovima

-
- ❑ Sortiranje relacija + linearni “prolaz” kroz relacije
 - ❑ Može se koristiti i heširanje ukoliko ga SUBP podržava
 - ❑ Pri implementaciji pristupnih rutina vrši se kombinacija operacija
-

Implementacija operacije SPOJ

□ Metod ugnježdenih petlji

- Najjednostavniji
 - Koristi se kad ne postoje pristupne strukture
 - Vreme izvršavanja proporcionalno ($n \times m$)
 - Gde su n i m brojevi n -torki u relacijama nad kojima se vrši operacija SPOJ-a
-

Implementacija operacije SPOJ

□ *Sort-Merge* metod

- Zahteva uređenost relacija po vrednosti obeležja spajanja
 - Kad su relacije unapred sortirane ima značajnu prednost u odnosu na metod ugnježenih petlji
 - Troškovi proporcionalni ($n \times \log m$)
-

Implementacija operacije SPOJ

□ Heš metod

- Jedan od najefikasnijih načina fizičke implementacije operacije SPOJ-a.
 - Performansa izvršavanja zavisi od *heš funkcije*.
 - Troškovi proporcionalni $(n + m)$.
-

Implementacija operacije SPOJ

❑ Korišćenjem posebnih struktura podataka

- Efikasne indeksne strukture
- Strukture za podršku samo fizičke implementacije operacije spoja:

Join indeksi	Indeksi koji se koriste i koriste samo za operaciju Spoj-a.
Kd –stablo	Višedimenzionalna samobalansirajuća B stabla
T – stablo	Vrsta binarnih stabala koja se koriste kada se indeksi i podaci skladište u operativnoj memoriji.

Zadatak optimizatora upita

- ❑ Određivanje alternativnih planova za izvršavanje upita
 - Radi smanjenja troškova optimizacije uglavnom se razmatra samo podskup svih mogućih planova
 - ❑ Estimacija troškova alternativnih planova i izbor plana sa najnižim troškovima
-

Osnovne tehnike optimizacije

☐ **Heuristička optimizacija** -

Heuristička optimizacija transformiše stablo upita korišćenjem skupa pravila koja tipično (ali ne u svim slučajevima) poboljšavaju performanse izvršenja.

☐ **Optimizacija estimacijom troškova izvršavanja upita** -

Optimizacija zasnovana na troškovima je "skupa", SUBP-ovi mogu koristiti heuristiku optimizaciju da smanje broj izbora koji se moraju napraviti estimacijom troškovima.

☐ **Semantička optimizacija** - Proces transformacije upita koji je izdao korisnik u drugi upit koji, zbog semantike daje isti odgovor.

Heuristička optimizacija

- ❑ Zasniva se na transformacionim pravilima;
 - ❑ Polazi se od kanoničkog stabla upita;
 - ❑ Vrš se transformacija polaznog, kanoničkog, stabla upita u konačno stablo upita koje ima bolju performansu;
 - ❑ Vrš se izbor pristupnih rutina i algoritama za operacije upita koje je moguće primeniti za raspoložive pristupne putanje baze podataka
 - ❑ Osnovno pravilo je unarne operacije izvršavati pre binarnih;
-

Optimizacija estimacijom troškova izvršavanja upita (1 / 2)

- ☐ Vrši se procena i poređenje troškova različitih planova izvršavanja upita;
 - ☐ Odabira se plan sa najnižim očekivanim troškovima;
 - ☐ Obično se ograničava broj planova koji se pri optimizaciji razmatra da bi se trošilo manje vremena;
 - ☐ Način optimizacije pogodan za kompilirane upite;
-

Optimizacija estimacijom troškova izvršavanja upita (2/2)

- ❑ Kod nekih SUBP se vrši “potpuna” optimizacija kompiliranih i “delimična” optimizacija interpretativnih upita, koja zahteva manje vremena;
 - ❑ Funkcija troškova je aproksimativna, pa je moguć i izbor plana izvršenja koji nije optimalan;
-

Ažuriranje relacionih baza podataka

- ❑ Neophodno je analizirati troškove pristupa podacima i troškove samog ažuriranja
 - ❑ Ograničavajući faktor na broj pristupnih struktura je odnos smanjenja troškova koje neki indeks donosi i troškova njegovog održavanja;
-

Ažuriranje relacionih baza podataka

- ❑ Troškovi ažuriranja n -torki i indeksa imaju važan uticaj na rezultat procesa selekcije indeksa i na odluke optimizatora;
 - ❑ U zavisnosti od toga da li se objektima BP pristupa istim redosledom kojim su smešteni u bazu podataka ili ne primenjuju se različite formule za računanje troškova;
-

Korisničko fizičko projektovanje BP

- ❑ Sledi posle projektovanja konceptualne šeme baze podataka;
 - ❑ Cilj je obezbediti dobru performansu često zahtevanih upita i operacija ažuriranja nad BP koja se projektuje;
-

Korisničko fizičko projektovanje BP

- ❑ Polazna osnova FP je logički opis podatataka koji je rezultat *logičkog projektovanja*;
 - ❑ Svaki SUBP podržava određeni skup fizičkih struktura podataka, pa se FP može definisati kao proces stvaranja efikasne fizičke strukture BP za konkretni SUBP, na osnovu konceptualne šeme;
-

Korisničko fizičko projektovanje BP

- ❑ Rezultat fizičkog projektovanja je *fizička šema* BP za konkretni (odabrani) SUBP.
 - ❑ Pri FP nije cilj samo doći do fizičke strukture BP, već to uraditi na način koji garantuje dobru performansu Baze podataka i celog IS.
 - ❑ Pošto se korisnički zahtevi proširuju i menjaju tokom vremena, da bi se obezbedila dobra performanse uglavnom je neophodno *podešavanje* parametara BP pri uvođenju sistema u rad, a kasnije tokom eksploatacije.
-

Zahtevi pri fizičkom projektovanju

Da bi se minimizirali troškovi U/I operacija u sistemu baze podataka važno je:

- a) da strukture na sekundarnoj memoriji obezbeđuju pretraživanje relevantnih podataka preko efikasnih pristupnih putanja;
 - b) da su podaci organizovani i smešteni na sekundarnu memoriju na način koji minimizira U/I troškove pri pristupu podacima.
-

Opcije pri fizičkom projektovanju

- ❑ Najniži nivo FP baza podataka obuhvata:
 - Izbor formata datoteka; definiše se broj datoteka BP, veličina datoteka, veličina stranica datoteka baze podataka, način grupisanja stranica u jedinice U/I prenosa, veličina kontinualnih proširenja ...
 - Korišćenje pokazivača; podržavaju navigaciju kroz BP; unapređuju performansu pri uparivanju skupova podataka; implementiraju se kao identifikatori n -torki (*IDN*-ovi)
-

Opcije pri fizičkom projektovanju

- ☐ Vertikalnu podelu relacija; rezultuje smanjenjem broja pristupa sekundarnoj memoriji; grupe obeležja iste relacije smeštaju se u različite datoteke BP.
 - ☐ Kompresiju podataka; od interesa je iz dva razloga: smanjenje potrebnog prostora na sekundarnoj memoriji i poboljšanja performanse obrade;
-

Opcije pri fizičkom projektovanju

- ❑ Asocijativno pretraživanje;
 - Koristi se da bi se smanjio broj pristupa sekundarnoj memoriji;
 - Postoji u svim SUBP; što je razlog da bez obzira na odabrani SUBP **izbor indeksa** predstavlja aktivnost FP koja se uvek sprovodi bilo da se radi o projektovanju jednostavnih ili složenih baza podataka;
 - Najpoznatija i najčešće korišćena indeksna struktura u relacionim bazama podataka je B-stablo (različite varijante – zavisno od konkretnog SUBP-a).
-

Asocijativno pretraživanje (nastavak)

- Mogu se koristiti tako da redosled i organizacija indeksa određuje redosled n -torki u datoteci podataka; *grupišući (klaster) indeksi*;
 - Ostali indeksi se nazivaju *negrupišući*; moraju biti gusti, odnosno postoji isti broj ulaznih tačaka u indeksnoj strukturi koliko ima n -torki;
 - Najveći broj SUBP pri pretraživanju indeksa ne pristupa n -torkama, pa je u nekim slučajevima do podataka moguće doći samo pretraživanjem indeksa
-

Opcije pri fizičkom projektovanju

- ❑ Kontrolisano uvođenje redudantnih podataka
 - Replikacija - opcija FP kojom se definiše čuvanje istih delova baze podataka na više uređaja sekundarne memorije;
 - Izvedene informacije - odnosno materijalizovani relacioni pogledi, nastaju kao posledica smeštanja rezultata operacije spoja na sekundarnu memoriju;
-

Opcije pri fizičkom projektovanju

- ❑ **Join indeksi** - U izvedene relacije se smeštaju samo neophodne informacije za efikasno izvršavanje operacije spoja; eliminišu problem ažuriranja;
 - ❑ **Denormalizacija** - Replikacija pojedinačnih vrednosti obeležja radi bržeg pristupa; značajna opcija za sisteme koji ne podržavaju grupisanje po različitim tipovima podataka;
-

Opcije pri fizičkom projektovanju

- ❑ Način smeštanja i fizički redosled podataka utiču na performansu pristupa podacima:
 - Grupisanje podataka - Organizacija elemenata podataka na sekundarnoj memoriji vrši se na takav način da se maksimizira količina relevantnih podataka koji se čitaju jednom U/I operacijom i na taj način smanji broj potrebnih U/I operacija da bi se zadovoljio zahtev obrade BP;
 - Razdvajanje relacija - Kod nekih SUBP i stranica datoteka BP po više uređaja sekundarne memorije omogućava veći broj U/I operacija u jedinici vremena, a time i veću U/I propustnost.
-

Završna razmatranja opcija FP

- ☐ Većina analiziranih opcija egzistira nezavisno od modela podataka;
 - ☐ Zbog velikog broja opcija FP predstavlja kompleksan zadatak;
 - ☐ Izvor kompleksnosti FP je i to što je većina odluka međuzavisne;
 - ☐ U komercijalnim SUBP nisu sve opcije FP podjednako zastupljene i nisu sve podjednako važne;
-

Izbor indeksa za tebele relacione BP

☐ Obuhvata:

- Izbor redosleda n -torki u relaciji.
- Izbor obeležja za definisanje skupa indeksa.

☐ Zavisi od:

- Karakteristika podataka.
 - Karakterističnog skupa aplikacija.
-

Karakteristike podataka

❑ Karakteristike podataka obuhvataju:

- Broj relacija;
 - Očekivani broj n -torki svake relacije;
 - Broj obeležja svake relacije;
 - Za svako obeležje:
 - Tip obeležja,
 - Dužina u bajtima,
 - Očekivani broj mogućih vrednosti koje obeležje dobija u n -torkama relacije;
 - veličina svake relacije izražena u broju osnovnih jedinica U/I prenosa.
-

Opis karakterističnog skupa aplikacija

- ❑ Spisak upita i frekvencije njihovog izvršavanja;
 - ❑ Spisak ažuriranja i njihove frekvencije izvršavanja;
 - ❑ Ciljnu performansu za svaki tip upita i ažuriranja.
-

Opis upita u skupu aplikacija

- ☐ Kojim relacijama se pristupa;
 - ☐ Koja obeležja se zahtevaju (u SELECT klauzuli);
 - ☐ Koja obeležja učestvuju u izrazima za uslove selekcije i spajanja relacija (u WHERE klauzuli) i očekivanu selektivnost tih uslova.
-

Opis ažuriranja u skupu aplikacija

- ❑ Nad kojim obeležjima je definisan uslov selekcije ili spajanja (u WHERE klauzuli) i očekivanu selektivnost tih uslova;
 - ❑ Tip ažuriranja (INSERT, DELETE ili UPDATE), i relacije koje se ažuriraju;
 - ❑ Za naredbu UPDATE, obeležja čije vrednosti se modifikuju.
-

Obeležja za definisanje pristupnih putanja

☐ Konstantni indeksi

☐ Pogodna obeležja

- Obeležje nad kojim postoji predikat u naredbi, i ako sistem može koristiti (neki) indeks za obradu tog predikata (WHERE klauzula)
 - Obeležja koja se pojavljuju u okviru GROUP BY i ORDER BY klauzula
-

Preporuke za selekciju indeksa

- Postupak selekcije
 - Kriterijumi za selekciju
-

Kriterijumi za selekciju indeksa

- Kada indeksirati;
 - Izbor ključa pretraživanja;
 - Složeni ključ pretraživanja;
 - Grupisanje;
 - Heširanje ili stabla indeksa;
 - Uravnoteženje troškova održavanja indeksa;
-

Indeksi nad Relacijama (Tabelama) Baze podataka (1/3)

- ❑ Indeksi su pomoćne strukture koje se kreiraju da bi se ubrzao pristup podacima u Relacijama (Tabelama) prilikom pretraživanja.
 - ❑ Kreiraju ih Administrator baze Podataka (ABP) preko SUBP-a.
 - ❑ SUBP ih automatski održava posle njihovog kreiranja.
 - ❑ Šta znači da se indeksna struktura (B-Stablo) drži balansiranim?
-

Indeksi nad Relacijama (Tabelama) Baze podataka (1/ 2)

- ❑ Sa aspekta grupisanja podataka postoji dve vrste indeksa:
 - Grupišući (*Cluster*) indeksi
 - Nad tabelom može biti samo jedan Grupišući (*Cluster*) indeks.
 - NeGrupišući (*Non Cluster*) indeksi
 - Nad tabelom može biti više Negrupišući (*NonCluster*) indeks-a.
-

Indeksi nad Relacijama (Tabelama) Baze podataka (3/ 3)

- ❑ Sa aspekta jedinstvenosti vrednosti indeksnog ključa postoji takođe dve vrste indeksa:
 - Jedinstveni indeksi
 - Nad jednom tabelom može biti definisano više jedinstvenih indeksa (nad različitim skupom obeležja).
 - Nejedinstveni indeksi
 - Nad tabelom takođe može biti definisano više nejedinstvenih indeksa.
 - ❑ Šta je indeksni ključ?
 - Može biti Prost i Složen.
-