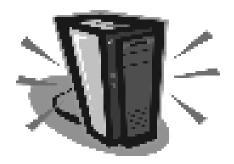
Podatkovne baze

Visokošolski študij

Bolonjski program



Matjaž Kukar, 2015

Splošne informacije...

Predavanja

Doc. dr. Matjaž Kukar
 <u>matjaz.kukar@fri.uni-lj.si</u>

– Govorilne ure: ???, kabinet R2.04

Vaje

Razpored skupin na spletni učilnici

Doc. dr. Luka Šajn
 luka.sajn@fri.uni-lj.si

Matej Pičulin
 matej.piculin@fri.uni-lj.si

Viri

- Prezentacije s predavanj/vaj niso dovolj!!!
- Glavna literatura
 - S. Sumathi, S. Esakkirajan: Fundamentals of Relational Database Management Systems, Springer, 2007
 - Paul Wilton and John W. Colby: Beginning SQL, Wrox, 2005.
 - A. Oppel (2004). Databases demystified, McGraw-Hill.
- Ostala priporočena literatura
 - [1] Raghu Ramakrishnan, Johannes Gehrke (2003). Database Management Systems, Third Edition, McGraw-Hill
 - [2] Thomas M. Connolly, Carolyn E. Begg (2005). Database Systems, A Practical Approach to Design, Implementation and Management, Fourth Edition, Addison-Wesley
 - [3] Ramez Elmasri, Shamkant B. Navathe (2003). Fundamentals of Database Systems, Fourth Edition, Addison-Wesley
 - [4] Tomaž Mohorič (2002). Podatkovne baze, Založba Bi-TIM.
 - [5] Peter Rob, Steven Morris, Carlos Coronel (2010). Database Systems: Design, Implementation and Management, 9th Edition, Addison Wesley.

Vsebina predavanj

- I. Osnove
 - Uvod v podatkovne baze: PB in SUPB
 - Podatkovni modeli
 - Relacijski podatkovni model
- II. Relacijsko poizvedovanje
 - Formalni poizvedovalni jeziki
 - <u>SQL</u>
- III. Sistemi za upravljanje s podatkovnimi bazami (SUPB)
 - Arhitektura
 - Koncepti
 - <u>Upravljanje</u>
- IV. Načrtovanje PB
 - Normalizacija
 - Konceptualno načrtovanje
- V. Napredne teme s področja PB: transakcije, varnost, ...
- VI. Nerelacijske podatkovne

Vsebina vaj

- I. Podrobnejša obravnava nekaterih tem s predavanj
 - Spoznavanje s programsko opremo
 - Relacijski podatkovni model in relacijska algebra
 - SQL (MariaDB, PostgreSQL)
 - Programski dostop do podatkovne baze
 - Indeksiranje
 - Normalizacija
- II. Domače naloge in kvizi (50% ocene vaj)
 - Pet (5) praktičnih domačih nalog s področja relacijske algebre in SQL
 - Oddaja preko spletne učilnice, približno na dva tedna
- III. Seminarska naloga (50% ocene vaj)
 - Na koncu semestra (april, maj)
 - Nekoliko obsežnejša kot domače naloge

Pravočasnost izdelave, oddaje in zagovora!

Praktično delo s PB

- FRI: SUPB na FRI, možen priklop iz učilnic in od drugod
 - MariaDB 10 (MySQL Workbench ali HeidiSQL client, spletni dostop)
 - PostgreSQL 9.4 (HeidiSQL client, spletni dostop)
 - (MongoDB kot primer nerelacijskega sistema)
 - Zakaj trije različni sistemi?
- Individualno delo
 - Zelo priporočljivo
 - MariaDB 10
 - MySQL workbench ali HeidiSQL
 - Povezave na učilnici

MariaDB, MySQL, PostgreSQL ???

- Vsi trije so odprtokodni sistemi
- MySQL: najpogosteje uporabljan sistem
- MariaDB ≅ MySQL
 - MySQL je pod Oraclovim okriljem; dvomi glede prihodnosti
 - raje uporabljamo kompatibilni, neodvisni MariaDB
- PostgreSQL: zahevnejši od MariaDB/MySQL, boljša podpora standardom (MySQL: spletne aplilkacije, PostgreSQL "resne" podatkovne aplikacije)
- Težišče dela bo na sistemu MariaDB, PostgreSQL pretežno za primerjavo

Izpitni red

- 1. Iz domačih nalog in seminarja morate doseči skupno najmanj polovico vseh možnih točk, da lahko pristopite k izpitu.
- 2. Sodelovanje na predavanjih in vajah se lahko nagradi (do 10% + 10% točk, subjektivna ocena)
- 3. Pisni izpit morate za pozitivno oceno pisati najmanj 50%, k čemer se potem prištejejo točke vaj. Po potrebi tudi ustni izpiti!
- 4. Pisni izpit lahko nadomestite z dvema neobveznima kolokvijema začetek aprila, začetek junija, vsak vsaj 50%
- 5. Veljavnost vaj: do vključno 30. 9. tekočega šolskega leta!

Izpit: Izpit: 72% Izpit: 49% Izpit: -

vsaj 50% Vaje: 83% Vaje: 83% Vaje: 49%

Skupaj: 77.5% Ocena: negativno

Vaje: Sodelovanje: 10% Ni izpolnjen pogoj

vsaj 50% Ocena: 9 za pristop k izpitu.

Alternativa: individualni projekt

- Če imate veliko predznanje s področja podatkovnih baz, lahko del obveznosti opravite z izvedbo in predstavitvijo individualnega projekta
 - v dogovoru z izvajalci
 - odvisno od obsega dela

Primer:

- predstavitev in uporaba sistema SQLite
- predstavitev in uporaba katerega od nerelacijskih sistemov za upravljanje s podatkovnimi bazami

Kaj bo treba zares dobro znati

- Poznavanje temeljnih konceptov relacijskih PB
- Samostojna vzpostavitev SUPB (MariaDB ali MySQL)
- Suverena uporaba orodij za delo s SUPB
- Suvereno programiranje v SQL
- Obvladovanje programskega dostopa do PB (Python, ODBC)
- Razumevanje in uporaba indeksiranja
- Osnove načrtovanja PB (normalizacija, ER)



Usposobljeni za samostojno in učinkovito delo s PB

Poglavje I **Uvod v podatkovne baze**Osnove

- Splošno o podatkovnih bazah
- Zgodovina shranjevanja podatkov
- Datotečni sistem in SUPB
- Opisovanje in shranjevanje podatkov v PB
- Poizvedovanje v PB
- Obvladovanje transakcij
- Zgradba SUPB

Splošno o podatkovnih bazah (PB)

Stanje danes:

- Organizacije odvisne od zmožnosti pridobivanja natančnih in pravočasnih podatkov...
- Podatki predstavljajo konkurenčno prednost.
- Brez zmožnosti za upravljanje z velikimi količinami podatkov in zmožnosti za hitro iskanje ustreznih podatkov postanejo podatki breme za organizacijo.
- Paradoks: zaradi preveč informacij potrebujemo še več informacij (informacije o informacijah)
- Potrebujemo ustrezne mehanizme za upravljanje s podatki in učinkovito iskanje po njih → Podatkovne baze.

Terminologija: ali ločimo med pojmoma?

- Podatkovna baza (PB)
- Sistem za upravljanje s podatkovnimi bazami (SUPB)

Definicija PB

 Podatkovna baza je avtomatizirana, večuporabniška, formalno definirana in (centralno) nadzorovana zbirka logično povezanih podatkov.

<u>Definicija SUPB</u>

- Sistem za upravljanje s podatkovno bazo SUPB (angleško: Database Management System – DBMS) je splošnonamenski skupek programske opreme, ki omogoča kreiranje, vzdrževanje in nadzor nad dostopom do podatkov v PB.
- Obstaja veliko vrst SUPB. Omejili se bomo predvsem na relacijske ter omenili sodobne nerelacijske:
 - Oracle, Sybase, DB2, MS SQL, Ingres, Postgres, MySQL
 - BigTable (Google), Dynamo (Amazon), Cassandra (Facebook, Twitter), Hadoop (Yahoo), Mongo DB, Couch DB, ...

Alternativa SUPB

- Shranjevanje v aplikaciji lastni obliki;
- Problemi: neprenosljivost, nefleksibilnost, redundantnost ...

- Namenske aplikacije: od začetka razvoja računalništva
- Zgodnja 60': Charles Bachman iz General Electric-a razvije prvi splošno-namenski SUPB (Integrated Data Store).
 - Predstavlja osnovo za mrežni podatkovni model, ki je predlagan za standard na konferenci za jezike podatkovnih sistemov (CODASYL).
 - Za mrežni podatkovni model Bachman prejme Turingovo nagrado (ACM Turing Award - na področju računalniških znanosti ekvivalent Nobelove nagrade)
 - Ima velik vpliv na razvoj SUPB-jev v 60' letih.
 - Zelo zapleteno vzdrževanje!

- Pozna 60': IBM trži Information Management
 System (IMS), ki se ponekod uporablja še danes.
 - Zasnovan na GUAM sistemu (NASA, Apollo Moon)
 - Predstavlja osnovo za hierarhični podatkovni model.
 - American Airlines in IBM razvijeta sistem SABRE za rezervacije letalskih kart – sistem omogoča več uporabnikom dostop do skupnih podatkov preko mreže.
 - Zanimivost: isti SABRE se še danes uporablja za spletna potovalna agencija Travelocity.

- 70' leta: Edgar Codd (IBM) predlaga relacijski podatkovni model.
 - Razvije se mnogo relacijskih SUPB.
 - Podatkovne baze postanejo akademsko področje. Razvije se izjemno močna teoretična podlaga.
 - Codd dobi Turingovo nagrado za svoje delo.
 - Relacijske podatkovne baze postanejo de-facto standard v podatkovnih sistemih.

- 80' leta: Relacijski model si še utrdi položaj kot SUPB.
 - Razvije se SQL poizvedovalni jezik (IBM-ov projekt System/R)
 - SQL postane standardni jezik za izvajanje poizvedb v relacijski PB.
 - SQL je bil standardiziran v poznih 80' SQL-89.
 - Standard prevzamejo American National Standard Institute (ANSI)
 in International Standards Organization (ISO).
 - Skrb za sočasen dostop do podatkov prevzame SUPB (transakcije in dostopni protokoli). Programerji programirajo, kot bi do podatkov dostopali samo oni. James Grey dobi za dosežke na tem področju Turingovo nagrado.

- Pozna 80' in 90': veliko napredka na področju PB.
 - Veliko raziskav se opravi na področju poizvedovalnih jezikov in bogatejših (razširjenih) podatkovnih modelov.
 - Velik poudarek na kompleksnih analizah podatkov iz vseh področij organizacijskih sistemov.
 - Veliko proizvajalcev SUPB-jev (IBM DB2, Oracle 8, Informix UDS) povzame ideje objektnega programiranja in razširi svoje sisteme s podporo novim podatkovnim tipom: slike, tekst in s podporo kompleksnejšim poizvedbam.

Pozna 80' in 90': nadaljevanje

- Pojavijo se podatkovna skladišča, ki združujejo podatke iz več PB in omogočajo izvajanje specializiranih analiz (iskanje zakonitosti v podatkih).
- Pojavijo se ERP (enterprise resource planning) in MRP (management resource planning) paketi
 - Podpirajo skupne funkcije v poslovnih sistemih (npr. skladiščno poslovanje, planiranje človeških virov, finančne analize,...).
 - Predstavljajo obsežno aplikacijsko plast nad skupno PB.
 - Primeri: Baan, Oracle, PeopleSoft, SAP in Siebel.

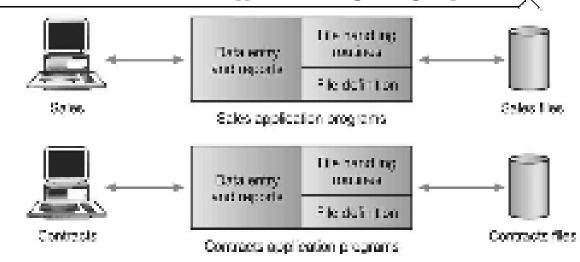
- Naslednja stopnja: vstop SUPB v svet Interneta
 - Prva generacija spletnih mest shranjuje podatke v datotekah
 OS. Uporaba PB za shranjevanje podatkov, ki so dostopni
 preko Interneta, postaja vsakdanja.
 - Poizvedbe se generira preko spletnih form, odgovore pa se nazaj posreduje v obliki jezika HTML, za lažji prikaz v spletnem brskalniku.
 - CMS, Blog, Wiki sistemi (hranjenje spletnih vsebin v PB)
 - Mnogouporabniške spletne in druge omrežne igre
 - Vsi proizvajalci dodajajo svojim SUPB-jem možnosti za čim lažjo uporabo v spletu.

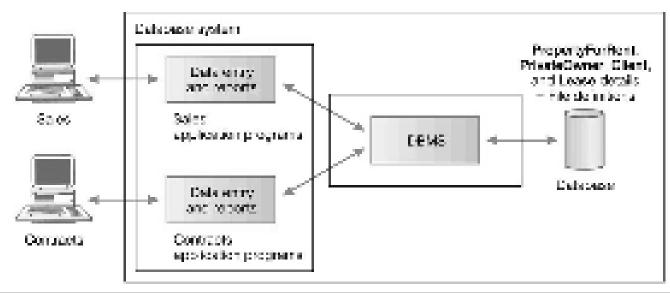
Najnovejša področja:

- Bogati podatkovni tipi (multimedija, prostorski podatki, XML), digitalne knjižnice,
- PB za interaktivni video, spletne skupnosti: masivna skalabilnost
- Odločitveni sistemi, vrtanje po podatkih, odkrivanje zakonitosti
- Raziskovalni projekti z intenzivno uporabo PB:
 - Human Genome Project, GEOSS The Global Earth Observation System of Systems (GEOSS), Raziskovanje vesolja
- "Webscale podatkovni" sistemi
 - "Big Data", podatkovni tokovi
 - Nerelacijske podatkovne baza (NoSQL)
 - NewSQL

- Shranjevanje podatkov v datotečnem sistemu:
 - zgodovinsko predhodnik SUPB
- Omejitve datotečnih sistemov:
 - Ločevanje in izolacija podatkov
 - Vsaka aplikacija obvladuje svoje podatke
 - V okviru ene aplikacije lahko dostopamo le do podatkov te aplikacije (ne vemo za razpoložljive podatke v drugih aplikacijah)
 - Podvajanje podatkov
 - Podvajanje vnosa iste podatke vnašamo večkrat in na več mest
 - Odvečna poraba prostora na disku ter potencialna možnost za neskladnost
 - Isti podatki podani v različnih oblikah

Datotečni sistemi in SUPB (podvajanje)





- Omejitve datotečnih sistemov (nadaljevanje):
 - Podatkovna odvisnost
 - Struktura podatkov je definirana v aplikaciji; sprememba v strukturi podatkov zahteva spremembo v aplikaciji
 - Neskladnost med oblikami datotek
 - Aplikacije, napisane v različnih programskih jezikih, ne morejo enostavno dostopati do datotek drugih sistemov

Datotečni sistemi in SUPB (podatkovna odvisnost)

	0	1	2	3	4	5	6	7	8	9
0	N	D	T	D	k		-	-		
10	-	-	-	-	-		-		-	-
20	J	D	n	Ė	X		-	-	-	
30	6	3	1	2	3	4	5	6-	0	-11
40	0	2	0	0	. 0	3	K	0	W	a
50	č	-		-			•	-	•	-
60								-		
70							N1	A	j	9.
80	-						-6	7	80	9
90	+	3	2.	1	0	-0	0	1	0	0
100	0	0		•		-		-		-

	0	1	2	3	4	5	6	7	8	9
0	-6	3	1	2	3	4	3	6	-6	3
10	7	-0	7	0	1	1	6	0	-6	2
20	0	1	0	ä	6	7	8	9	4	3
30	2	1	6	3	×	0	7	0	1	- ()
40	2	0	9	2	0	_ 1	0	6	3	. 1
50	2	3	4	5	-6	6	3	7	0	7
60	0	2	0	0	2	0	9	2	0	1
70	-11			-						-
80				-						-
90			-	-	•		٠			-
100			-	-	•			•		-

- a. Datoteka s podatki o vpisanih študentih
- b. Datoteka z ocenami opravljenih polaganj izpitov

 Zaradi težav oziroma neučinkovitosti shranjevanja podatkov neposredno v datoteke, se pojavijo

sistemi za upravljanje s podatkovnimi bazami – SUPB

ki probleme datotečnega sistema rešujejo kot vmesni člen z vpeljavo podatkovnega modela in drugih tehnik.

SUPB je splošnonamenski skupek programske opreme, ki omogoča kreiranje, vzdrževanje in nadzor nad dostopom do podatkov v PB.

- SUPB uporablja različne mehanizme za upravljanje s podatki:
 - Kreiranje podatkovnih struktur je omogočeno z mehanizmi
 DDL Data Definition Language.
 - Omogoča definiranje podatkovnih struktur in tipov ter omejitev
 - Vse specifikacije so shranjene v PB (podatkovni slovar, sistemski katalog). Kreiranje tabele pomeni poseg v sistemski katalog.
 - Vzdrževanje podatkov (Create, Insert, Update, Delete)
 izvajamo z uporabo mehanizmov DML Data Manipulation
 Language.
 - Za izvajanje povpraševanja uporabljamo povpraševalne jezike (mehanizmi DQL - Data Query Language)

- SUPB zagotavlja mehanizme za nadzor nad dostopom do podatkov:
 - Varnost: dostop do podatkov v skladu z avtorizacijo
 - Skladnost: zagotavlja skladnost podatkov
 - Sočasen dostop: zagotavlja in nadzira sočasen dostop več uporabnikov do istih ali različnih podatkov
 - Obnova: zagotavlja mehanizme za obnovo podatkov

Uporaba SUPB prinaša naslednje prednosti:

- Podatkovna neodvisnost: Programi so neodvisni od predstavitve podatkov in načina shranjevanja podatkov.
 SUPB zagotavlja abstrakcijo podatkov in ločuje programe od podrobnosti predstavitve podatkov.
- Učinkovit dostop do podatkov: SUPB zagotavlja tehnike za učinkovito hranjenje in dostop do podatkov.
- Varnost in integriteta podatkov: Če se do podatkov dostopa preko SUPB, se lahko uporabi omejitve, ki zagotavljajo skladnost podatkov.

Prednosti uporabe SUPB (nadaljevanje):

- Administracija podatkov: Če so podatki shranjeni centralno, je upravljanje s podatki lažje.
- Sočasen dostop do podatkov in obnavljanje PB: SUPB razporeja sočasne dostope tako, da izgleda, kot da do podatkov dostopa en uporabnik.
- Skrajša čas razvoja programov: SUPB podpira številne mehanizme in orodja za dostop do podatkov.

- Zaključek: SUPB prinaša veliko prednosti v primerjavi s hranjenjem podatkov neposredno v datotekah.
- Obstajajo (redke) izjeme, ko uporaba SUPB ni primerna. Npr.:
 - Za specializirane aplikacije (npr. v realnem času) klasični SUPB niso vedno primerni. Za te aplikacije se raje izdela namensko kodo za rokovanje s podatki ali uporabi NoSQL.
 - Če SUPB ne podpira dela s podatki na način, ki ga zahteva program.

Sistem za upravljanje s pod. bazami

Komponente SUPB

- Strojna oprema
- Programska oprema



Postopki

Ljudje



Logično povezani

podatki znotraj PB

Sistem za upravljanje s pod. bazami

Komponente SUPB

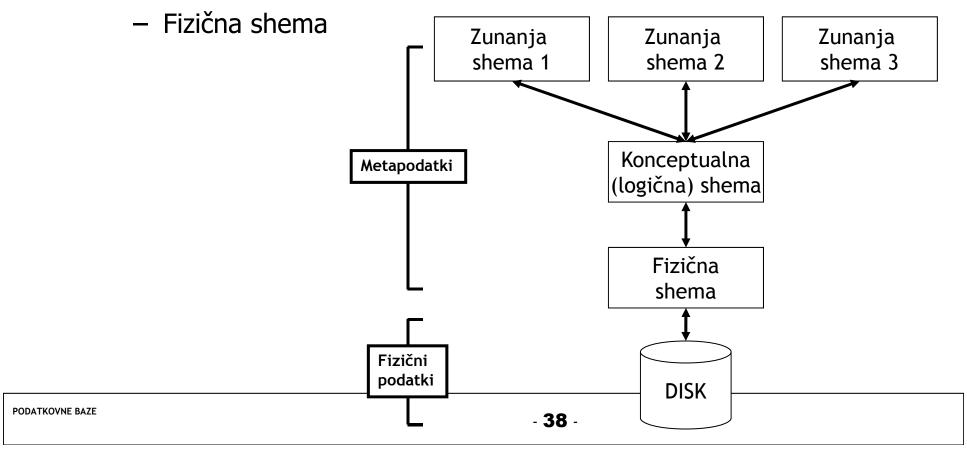
- Strojna oprema:
 - Podatkovni strežnik
 - Pomembna parametra za strežnik: hitri pomnilnik in diskovni prostor
 - Ostala strojna oprema (omrežje, UPS, ...)
- Programska oprema:
 - SUPB, operacijski sistem, omrežna programska oprema
 - Različna sistemska oprema (razvojna orodja, orodja za dostop do podatkov)
 - Specializirane aplikacije

Sistem za upravljanje s pod. bazami

- Komponente SUPB (nadaljevanje):
 - Postopki
 - Načini prijave
 - Uporaba posameznih orodij
 - Zagon in zaustavitev podatkovne baze
 - Izdelava varnostnih kopij
 - Obvladovanje nesreč/okvar
 - Ljudje, njihove vloge
 - Skrbnik podatkov
 - Skrbnik podatkovne baze (DBA Database Administrator)
 - Analitik, načrtovalec PB
 - Razvijalci aplikacij
 - Končni uporabniki (izkušeni, neizkušeni)

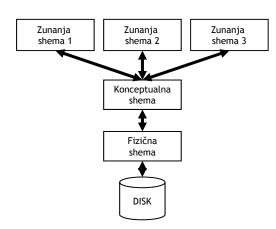
Tri-nivojska predstavitev podatkov (ANSI-SPARC)

- Podatki so v PB opisani na treh ravneh:
 - Zunanja shema
 - Konceptualna ali logična shema



Tri-nivojska predstavitev podatkov...

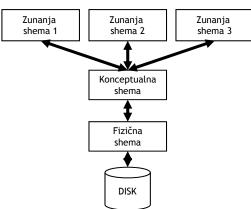
- Konceptualna ali logična shema
 - Konceptualna shema opisuje podatke z vidika podatkovnega modela, ki ga PB uporablja. Npr.:
 - Podatki o entitetah (profesor, študent, predavalnica,...)
 - Podatki o povezavah (predava, posluša,...).
 - Proces izdelave konceptualne sheme se imenuje konceptualno in logično načrtovanje.



Tri-nivojska predstavitev podatkov...

Fizična shema:

- Fizična shema podaja podrobnosti o shranjevanju podatkov
 → kako so podatki iz konceptualne sheme dejansko shranjene na sekundarnem pomnilniku: trdi disk, magnetni trakovi,...
- Odločiti se je potrebno, kakšno datotečno organizacijo bomo uporabili za shranjevanje podatkov in kreirati indeksne datoteke.
- Proces izdelave fizične sheme se imenuje načrtovanje fizične PB.



Tri-nivojska predstavitev podatkov

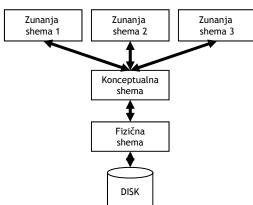
Zunanja shema:

- Tudi zunanje sheme uporabljajo koncepte podatkovnega modela (gradnike konceptualne sheme).
- Zunanja shema se uporablja za dostop do podatkov, ki je prilagojen določenemu uporabniku ali skupini uporabnikov.
- Vsaka zunanja shema sestoji iz enega ali več pogledov (view) in entitet iz konceptualne sheme.

Pogled je logična tabela, ki ne obstaja v fizični podatkovni

bazi.

 Izdelava zunanjih shem se izvaja v sodelovanju z uporabniki – zajemanje uporabniških zahtev.



Podatkovna neodvisnost...

- Tri-nivojska predstavitev podatkov v SUPB omogoča podatkovno neodvisnost - programi so neodvisni od načina shranjevanja in strukturiranja podatkov v PB:
 - Če se spremeni konceptualna shema, lahko zunanjo shemo priredimo tako, da pogledi ostanejo nespremenjeni -> logična podatkovna neodvisnost.
 - Podobno konceptualna shema ločuje uporabnike in načrtovalce od sprememb, ki se naredijo na fizični PB -> fizična podatkovna neodvisnost.

Podatkovna neodvisnost

- Konceptualna shema skrije podrobnosti o tem, kako so podatki dejansko shranjeni na disku, o strukturi datotek in o pomožnih dostopnih strukturah (npr. indeksih).
- Dokler ostaja konceptualna shema nespremenjena, spremembe na fizičnem nivoju ne vplivajo na programe, ki podatke uporabljajo.
- Lahko pa spremembe vplivajo na učinkovitost.

Poizvedovanje v PB...

- Enostavnost pridobivanja informacij iz PB je ključna prednost SUPB za uporabnike.
- Sodobni SUPB omogočajo uporabnikom postavljati enostavna vprašanja (poizvedbe), s katerimi pridobivajo podatke in iz njih izluščijo informacije.
- Poizvedbe se podaja v jeziku, ki je prirejen za opisovanje poizvedb – poizvedovalni jezik.
- Predvsem relacijski SUPB podpirajo zelo močne poizvedovalne jezike
- Ena pomembnih nalog SUPB je optimizacija poizvedb, tako da se te čim hitreje izvedejo.

Poizvedovanje v PB

- Učinkovitost poizvedb je močno odvisna od načina, kako so podatki shranjeni v fizični obliki.
- SUPB omogoča uporabnikom izvajati poizvedbe ter kreirati in posodabljati vrednosti s pomočjo DQL - Data Query Language in DML - Data Manipulation Language ukazov
- DML ukazi omogočjo vstavljanje, brisanje in spreminjanje zapisov v tabelah.

Poizvedovanje v relacijskih SUPB

- Temelji na relacijski algebri in relacijskem računu
- Morda malo dolgočasna, vendar nujna tematika
- Analogija: kako bi programirali v Pythonu, če ne bi vedeli kaj pomenijo osnovne aritmetične operacije (+-*/=)?
- Pustimo za kasneje, začnemo s preprostimi primeri