

P270 Релационе базе података

Ненад Митић

Математички факултет
nenad.mitic@matf.bg.ac.rs

Курсеви који чине целину

- 1 Релационе базе података
- 2 Програмирање база података
- 3 Пројектовање база података

Знања из РБП и ПБП су потребна у предметима

- 1 Истраживање података 1
- 2 Истраживање података 2
- 3 Курсевима на мастер студијама
- 4 ...

Садржај курса

- Основни појмови. Увод у базе података и упитни језик SQL; историја и мотивација
- Системи за управљање базама података; функција, архитектура; независност података
- Релациони модел података. Релациони упитни језици. DDL, DML, DCL компоненте SQL језика
- SELECT наредба. Кориснички дефинисане функције
- Релациони модел података - релациона алгебра и релациони рачун
- Погледи. Окидачи. MQT. Рекурзивни SQL. OLAP
- Интегритет и безбедност података у релационим базама података. Кључеви

Садржај курса

- Сигурност података и ауторизација
- Функционалне и вишезначне зависности. Нормалне форме и нормализација
- Увод у XML, XPath, XQuery и SQL/XML
- Трансакциона обрада података. Конкурентност и опоравак. Логови. Алгоритми опоравка
- Оптимизација у релационим базама података
- Администрација база података

Раст количине података

- Брз раст количине података
 - Свака два дана се количина података удвостручује
 - Дневно се формира више од *3ЕВ* нових информације - више него у комплетној историји цивилизације све до 2003. године
- Велике базе података
 - Нпр. База података о клими у Макс Планк институту за метеорологију (2001) \geq 220ТВ доступних на Веб-у + 6РВ додатних података
 - Подаци се ажурирају сваких 10 минута

Раст количине података - наставак

- Количина података се дуплира сваких 6 месеци
- \implies количина простора на дисковима и број продатих дискова се дуплирају у истом периоду
- Паркинсонов закон (модификација): подаци теже да испуне сав слободан простор у меморији
- Модификација Муровог закона: време потребно за обраду података се дуплира сваких 18 месеци

Раст количине података – наставак

Пошто се могућности корисника не повећавају истом брзином

⇒ потребне су ефикасније технике за чување и обраду података

Шта је систем база података

СБП (Систем База Података) је у основи систем за рачунарско записивање и чување слогова, тј. систем чија је сврха да чува информације и дозволи кориснику да те информације добије и ажурира по жељи

Главне компоненте СБП

- 1 Подаци
- 2 Хардвер
- 3 Софтвер
- 4 Корисници

Компоненте СБП - подаци

Подаци

- Интегрисани
- Дељиви

Компоненте СБП - хардвер

Хардвер

- Спољашњи меморијски уређаји
- Процесор и главна меморија

Компоненте СБП - софтвер

Софтвер

- СУБП (енг. *Database management system, DBMS*)
 - ниво софтвера који се налази између корисника и физичких података у бази
 - штити кориснике база од детаља на хардверском нивоу
 - управља свим захтевима за директан приступ бази
- Алати за развој апликација, писање извештаја, помоћни (*utility*) програми, програми за управљање трансакцијама (*TP monitor*)

Компоненте СБП – корисници

Корисници

- Апликативни програмери
- Крајњи корисници
- Администратори
 - Администратор базе података
 - Администратор података

Шта је база података?

- База података је скуп постојаних података који се користе од стране апликативних програма у неком окружењу
- Постојани подаци
 - када се једном нађу у бази не могу да буду уклоњени из базе без експлицитног захтева СУБП

Ентитети и односи

- Ентитети и односи
 - Ентитет: особа, место, ствар,
 - Однос: опис везе између два или више ентитета
 - Ако се ентитет дефинише само као објекат о коме желимо да сакупљамо информације, тада односи задовољавају ову дефиницију

Модел података

- Подаци
- Модел података је апстрактна, самостална, дефиниција објеката, оператора,, који заједно чине апстрактну машину са којом корисник комуницира
 - Објекти допуштају моделирање *структуре* података
 - Оператори допуштају моделирање *понашања*
- Имплементација датог модела је физичка реализација на стварној машини компонената апстрактне машине које заједно чине модел

Зашто база податка?

- Компактност
- Брзина
- Мањи напор
- Актуелност података
- Централизована контрола (у вишекорисничком окружењу)

Предности рада са базом

- Подаци могу бити дељени
- Смањење редувантности података
- Избегавање неконсистентности
- Подршка за трансакциони рад
- Одржавање интегритета
- Примена заштите података
- Балансирање између конфликтних захтева
- Примена стандарда

Администратор базе података

- Администратор базе података (енг. *database administrator, DBA*)
 - професионалац у ИТ
 - формира базу и имплементира контролне структуре
 - одговоран за имплементацију одлука DA
 - одговоран за рад система, перформансе, ...

Администратор података

- Администратор података (енг. *data administrator, DA*)
 - разуме постојеће податке
 - одлучује који подаци ће бити чувани у бази
 - установљава правила за одржавање и рад са подацима по њиховом чувању у бази
 - није техничко лице, већ припада управљачким структурама

Независност података

- Апликације имплементиране на старим системима су биле *зависне од података*
- Није погодно да апликације буду зависне од података у бази
 - различите апликације захтевају раличите погледе над истим подацима
 - ДВА мора да има слободу да промени физичку репрезентацију или приступне технике ради повећања перформанси

Независност података – наставак

- Дефиниција: независност података је отпорност апликације на промене физичке репрезентације података и приступних техника
- На које промене ДВА у бази апликација треба да буде отпорна?

Независност података – наставак

- Појмови

- сачувано поље (енг. *stored field*) је најмања јединица података која може да се чува
- сачувани слог (енг. *stored record*) је скуп сачуваних поља
- сачувана датотека (енг. *stored file*) је скуп свих тренутно постојећих сачуваних слогова истог типа

Независност података – наставак

- Аспекти сачуваних репрезентација који могу да буду предмет промена од стране DBA
 - репрезентација бројчаних података
 - репрезентација знаковних података
 - јединице за бројчане податке
 - кодирање података

Независност података - наставак

- База треба да буде способна да се шири без промене постојећих апликација
 - материјализација података
 - структура сачуваних слогова
 - структура сачуваних датотека
- База треба да буде способна да се шири без негативног утицаја на постојеће апликације

Ране 60-те године

- Charles Bachman (Тјурингова награда 1973.)
- Integrated Data Store
- Мрежни модел података
- CODASYL / DBTG
 - CODASYL - Conference on Data Systems Languages
 - DBTG - Data Base Task Group

Касне 60-те и 70-те године

- Касне 60-те
 - IBM
 - Information Management Systems (IMS)
 - Хијерархијски модел података
- 70-те
 - Edgar Codd, IBM (Тјурингова награда 1981.)
 - Релациони модел података

80-те године

- Доминација релационог модела
- SQL
- Управљање трансакцијама (James Gray, Тјурингова награда 1999.)

90-те године и данас

- Објектно-оријентисани модел података
- Објектно-релациони модел података
- Складиштење података (енг. *Data warehousing*) и истраживање података (енг. *Data mining*)
- Приступ базама података преко веба/интернета
- Мултимедијални и текстуални подаци (енг. *Information retrieval*)
- Структура података(XML)
- JSON
- NOSQL baze