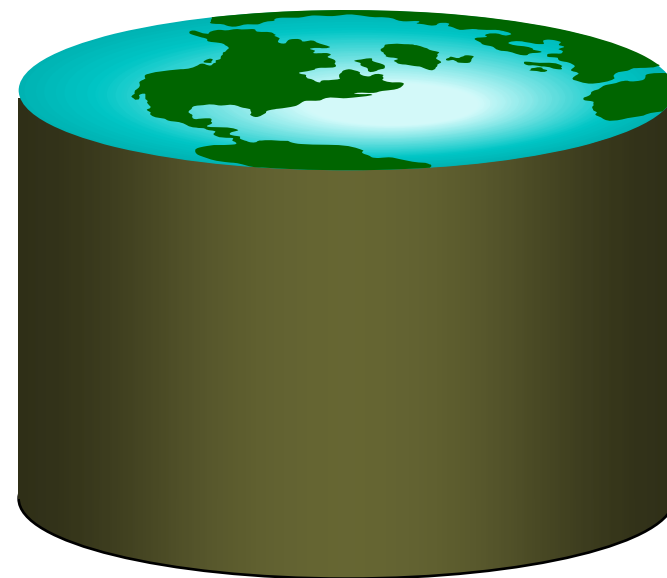




Baze Podataka

Miroslav Bojović

Jesen 2006.





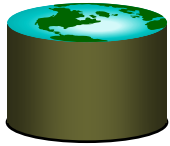
Pitanja za danas

- Ko?
- Kako?
- Šta?
- Zašto?



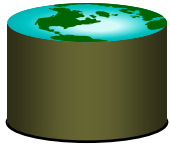
Ko?

- Nastavnik
 - Prof. Miroslav Bojović mbojovic@etf.bg.ac.yu
- Asistent
 - Miloš Cvetanović cmilos@etf.bg.ac.yu
- Osnovna literatura
 - *Database System Concepts*,
A. Silberschatz, H. Korth, S Sudarshan, McGraw Hill International Edition, 2005.
 - *Database Systems: The Complete Book*,
H. Garcia-Molina, J.D.Ulman, J. Widom, Prentice Hall, 2002.
 - Introduction to Database Systems,
Joe Hellerstein, UC Berkeley, M. Garofalakis, Intel Research, 2005.
 - *Upravljanje Transakcijama*,
M.Bojovic, Univerzitetska Misao, 2003.



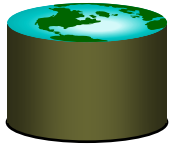
Kako?

- Lab vežbe: Projekat sa "real world" fokusom:
 - Microsoft SQL Server
- Ispit – Kolokvijum (30%+10%) i Finalni (70%)



Šta: Database Sistemi Nekada





Šta: Database Sistemi Danas

The screenshot shows the match.com homepage. A red circle highlights the Ask Jeeves search bar and the 'Search' button. Another red circle highlights the 'Members sign in here' link. A third red circle highlights the registration form, which includes the text 'Start browsing now. It's free.' and fields for gender, age, and location, with a 'Browse' button.

match.com

How it works | Success stories | Subscribe Now

Love is complicated. match is simple.

Start browsing now. It's free.

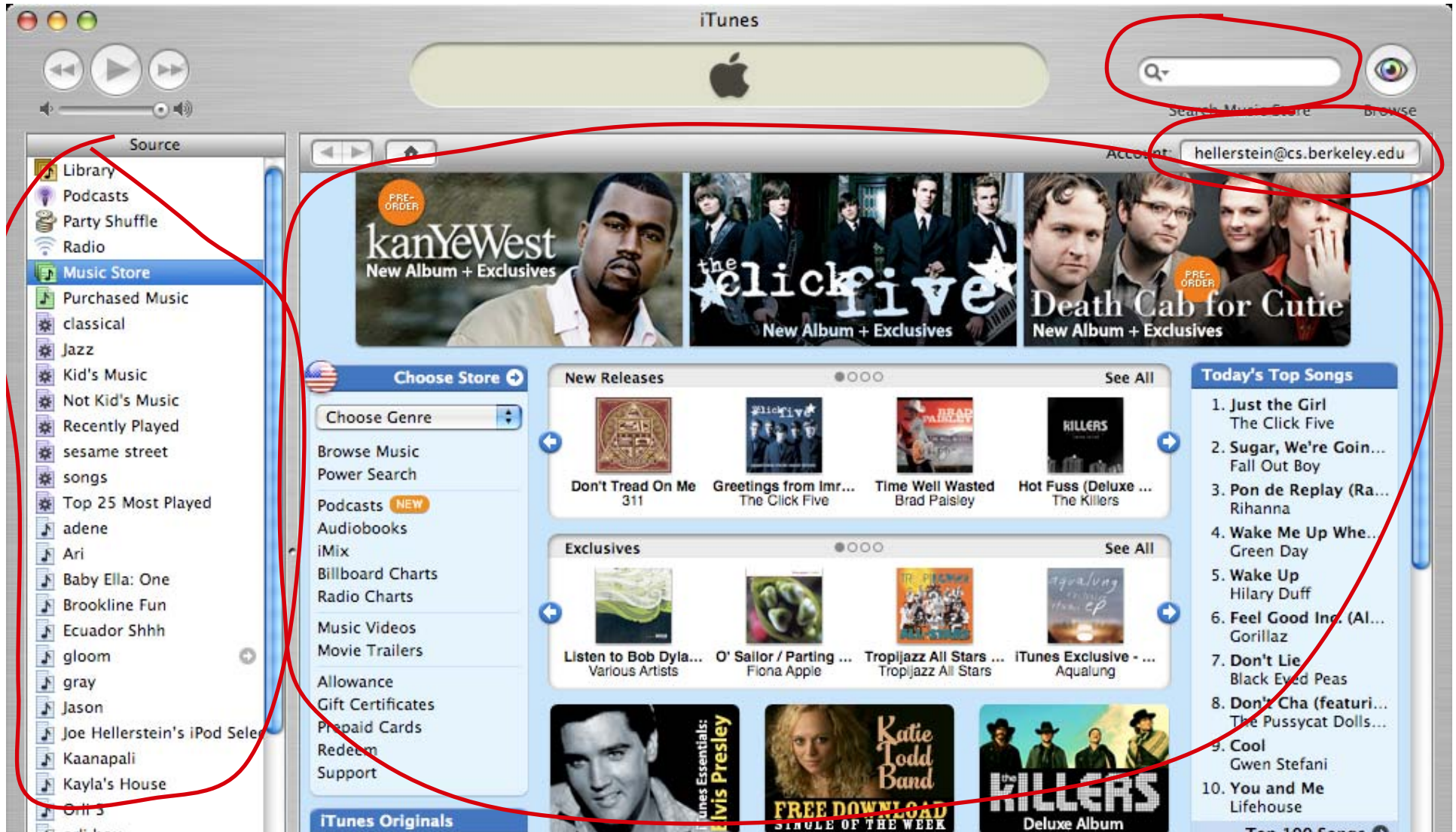
I am a seeking between and

located in (city/zip code)

Watch our new TV ads >>




Šta: Database Sistemi Danas





Šta: Database Sistemi Danas

**Higher Standards**

Accounts | **Bill Pay & e-Bills** | **Transfer Funds** | **Customer Ser**

Accounts Overview | **Account Activity** | Account Summary | Search

John Jones - Personal Accounts
Monday, January 12, 2004

I want to...
[View my account details](#)
[Set up a bill payment](#)
[Pay a bill](#)
[Transfer funds between accounts](#)

Announcements

Account
[Interest Checking - 3858](#)
[Regular Savings - 0490](#)
[Fixed Term CD -2747](#)
[Fixed Term IRA - 4128](#)



Šta: Database Sistemi Danas

NCBI

PubMed Nucleotide Protein Genome Gene Structure PopSet Taxonomy Help

Search for on chromosome(s) assembly All

☐ Show related entries

Entrez Genomes

[MapViewer Home](#) [Homo sapiens genome view](#) [BLAST search the human genome](#)

[Build 35.1 statistics](#)

Map Viewer
Help
Human Maps Help
Mouse Maps Help
NCBI Handbook

Related Resources
Human Genome Guide
Genomic Biology
Gene
OMIM
UniGene

Sequence Data
Human Genome
Sequencing
Mouse Genome
Sequencing
RefSeq

1 2 3 4 5 6 7 8 9 10 11 12 13

14 15 16 17 18 19 20 21 22 X Y

Lineage: [Eukaryota](#); [Metazoa](#); [Chordata](#); [Craniata](#); [Vertebrata](#); [Euteleostomi](#); [Mammalia](#); [Eutheria](#); [Euarchontoglires](#); [Primates](#); [Catarrhini](#); [Hominidae](#); [Homo](#); [Homo sapiens](#)



Prema tome... Šta je Baza podataka?

- A Database:
 - A very large, integrated collection of data.
- Typically models a real-world “enterprise”
 - Entities (e.g., teams, games)
 - Relationships (e.g. The A’s are playing in the World Series)
- Might surprise you how flexible this is
 - Web search:
 - Entities: words, documents
 - Relationships: word in document, document links to document.
 - P2P filesharing:
 - Entities: words, filenames, hosts
 - Relationships: word in filename, file available at host



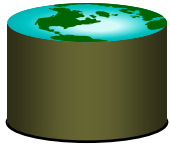
Šta je Sistem za Upravljanje Bazama Podataka(DataBase Management System)?

- DBMS contains information about a particular enterprise
 - Collection of interrelated data
 - Set of programs to access the data
 - An environment that is both *convenient* and *efficient* to use
- Database Applications:
 - Banking: all transactions
 - Airlines: reservations, schedules
 - Universities: registration, grades
 - Sales: customers, products, purchases
 - Online retailers: order tracking, customized recommendations
 - Manufacturing: production, inventory, orders, supply chain
 - Human resources: employee records, salaries, tax deductions
- Databases touch all aspects of our lives



Šta: Da li je WWW DBMS?

- Fairly sophisticated search available
 - Crawler **indexes** pages on the web
 - Keyword-based **search** for pages
- But, currently
 - data is mostly **unstructured** and **untyped**
 - **search only**:
 - can't modify the data
 - can't get summaries, complex combinations of data
 - **few guarantees** provided for freshness of data, consistency across data items, fault tolerance, ...
 - Web sites typically have a (relational) DBMS in the background to provide these functions.
- The picture is changing
 - New standards e.g., XML, Semantic Web can help data modeling
 - Research on combining/summarizing data across documents



Šta: "Search" vs. Query

- What if you wanted to find out which actors donated to John Kerry's presidential campaign?
- Try "actors donated to john kerry" in your favorite search engine.

Google Web Images Groups News Froogle Local more »

actors donated to john kerry Search Advanced Search Preferences

Web Results 1 - 10 of about 50,400 for [actors](#) [donated](#) to [john kerry](#) with Safesearch on. (0

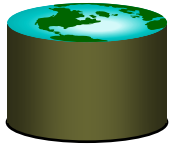
[The Forest For The Trees: The Gift That Keeps On Giving](#)
John Kerry donated \$25000 to the DNC but none to himself. ... Ted Danson, Actor, donated \$7000 to the DNC, \$2000 to Howard Dean, \$2000 to John Edwards and ...
[forester.blogspot.com/2004/07/gift-that-keeps-on-giving.html](#) - 24k - [Cached](#) - [Similar pages](#)

[Kerry's Donation Records Under Fire. Money From Chinese Army](#)
China-gate figure John Huang also donated money to Kerry. Huang later pled guilty to illegal fundraising for the DNC. Huang remains free despite the fact ...
[www.newsmax.com/archives/articles/2004/2/10/135320.shtml](#) - 35k - [Cached](#) - [Similar pages](#)

[Arnold a Rare Republican in Hollywood: Academy Award winners who ...](#)
... Oscar-winning actors, actresses and directors donated about 40 times more money to ... John Kerry, D-Mass., has raised more money from the entertainment ...
[www.isteve.com/2003_Few_Republicans_in_Hollywood.htm](#) - 15k - [Cached](#) - [Similar pages](#)

[Some stars have GOP stripes](#)
But, as the GOP showed when it invited wrestler-turned-actor The Rock to address ... "For someone like John Kerry, who doesn't have a lot of star presence ...
[www.fortliberty.org/forum/about73.html](#) - 31k - [Cached](#) - [Similar pages](#)

[\[PDF\] The Hollywood Left: John Kerry's Most Loyal Constituency](#)
File Format: PDF/Adobe Acrobat - [View as HTML](#)
lywood actors gave money to Kerry; only 12 gave to Bush. The GOP can ... Alec (Alexander) Baldwin \$2000 to John Kerry for President. • William Baldwin ...
[www.capitalresearch.org/pubs/pdf/FW0904.pdf](#) - [Similar pages](#)



Šta: "Search" vs. Query II

- What if you wanted to find out which **musicians** donated to John Kerry's presidential campaign?
- Try "**musicians donated to john kerry**" in your favorite search engine.
- If it isn't "published", it can't be searched

Google Web Images Groups News Froogle Local more »

musicians donated to john kerry Search [Advance](#) [Prefere](#)

Web Results 1 - 10 of about 22,300 for [musicians donated to john kerry](#) with Safese

[Tara Angell](#)
100% of every ticket sale is directly contributed to **John Kerry** for President.
... "These **musicians** are united by a want to change our country's leadership ...
[www.taraangell.com/archive.html](#) - 14k - [Cached](#) - [Similar pages](#)

[MoveOn.org: Democracy in Action](#)
[moveonpac.org/archive/](#) - [Similar pages](#)

[KyndMusic - it's the jam that matters at kyndmusic.com](#)
Music at the time was just music; and **musicians**, purely **musicians** – often scorned by
... prancing around the stage in a white suit and **John Kerry** t-shirt, ...
[www.kyndmusic.com/monthlyissues/n104/vfc.htm](#) - 21k - [Cached](#) - [Similar pages](#)

[Rock Stars Against Bush - A List of over 200 Bands Against Bush ...](#)
John Fogerty - Played fundraisers with "**Musicians for Kerry**". ... Don Henley -
donated money and time to the **Kerry** campaign. Played for Al Gore Benefit in ...
[www.spacechase.net/green/rock_against_bush.htm](#) - 78k - [Cached](#) - [Similar pages](#)

[Retro vs. Metro : Blog : Sep 22, 2004 - Sep 14, 2004](#)
If you ask Fox "News" viewers how they would characterize **John Kerry**, ...
Together, Lay and Enron **donated** \$2 million to Bush's 2000 campaign. ...
[www.retrovsmetro.org/blog/index/235](#) - 56k - [Cached](#) - [Similar pages](#)

[Political Topics And Discussion > Us Musicians Begin Anti-bush Tour](#)
A group of top US **musicians**, including Bruce Springsteen and REM, ... candidate **John Kerry**, before visiting the equally critical Florida on 8 October. ...
[www.bearpit.net/lofiversion/index.php/t2563.html](#) - 12k - [Cached](#) - [Similar pages](#)

[This, That, and The Other](#)



Šta: Da li je File Sistem DBMS?

- Thought Experiment 1:
 - You and your project partner are editing the same file.
 - You both save it at the same time.
 - Whose changes survive?

A) Yours B) Partner's C) Both D) Neither E) ???

- Thought Experiment 2:
 - You're updating a file.
 - The power goes out.
 - Which changes survive?

Q: How do you write programs over a subsystem when it promises you only "???" ?

A: Very, very carefully!!

A) All B) None C) All Since Last Save D) ???



Podrška OS za Upravljanje Podacima (Data Management)

- Data can be stored in RAM
 - this is what every programming language offers!
 - RAM is fast, and random access
 - Isn't this heaven?
- Every OS includes a File System
 - manages *files* on a magnetic disk
 - allows *open, read, seek, close* on a file
 - allows protections to be set on a file
 - drawbacks relative to RAM?



Da li Fajl Sistem zadovoljava? (1)

- In the early days, database applications were built directly on top of file systems
- Drawbacks of using file systems to store data:
 - Data redundancy and inconsistency
 - Multiple file formats, duplication of information in different files
 - Difficulty in accessing data
 - Need to write a new program to carry out each new task
 - Data isolation — multiple files and formats
 - Integrity problems
 - Integrity constraints (e.g. account balance > 0) become “buried” in program code rather than being stated explicitly
 - Hard to add new constraints or change existing ones



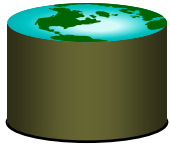
Da li Fajl Sistem zadovoljava? (2)

- Drawbacks of using file systems (cont.)
 - Atomicity of updates
 - Failures may leave database in an inconsistent state with partial updates carried out
 - Example: Transfer of funds from one account to another should either complete or not happen at all
 - Concurrent access by multiple users
 - Concurrent accesses needed for performance
 - Uncontrolled concurrent accesses can lead to inconsistencies
 - Example: Two people reading a balance and updating it at the same time
 - Security problems
 - Hard to provide user access to some, but not all, data
- Database systems offer solutions to all the above problems



Savremeni DBMS obezbeđuju:

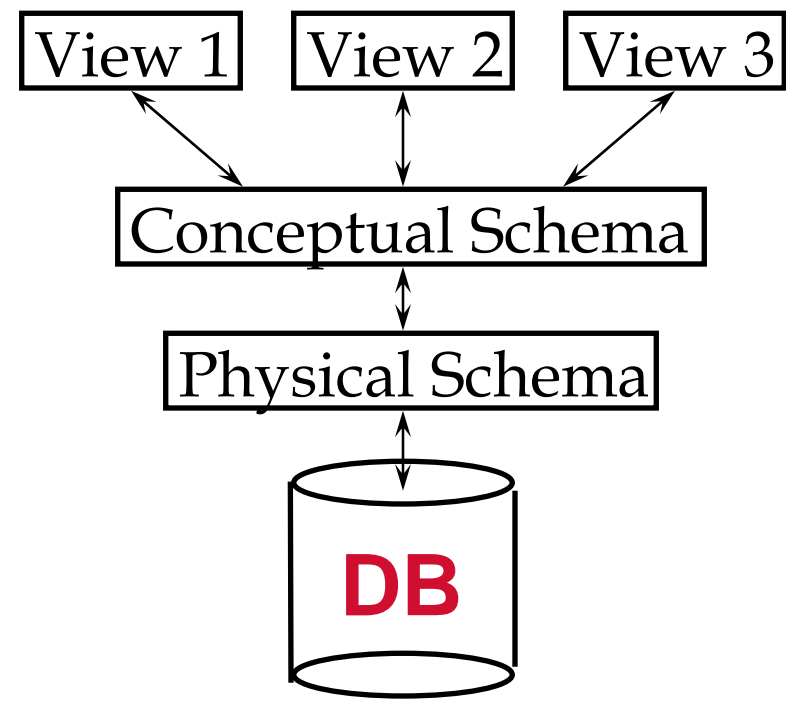
- Data independence
- Efficient data access
- Data integrity & security
- Data administration
- Concurrent access, crash recovery
- Reduced application development time
- So why not use them always?
 - Expensive/complicated to set up & maintain
 - This cost & complexity must be offset by need
 - General-purpose, not suited for special-purpose tasks (e.g. text search!)



Nivoi Apstrakcije Podataka

- **Pogledi (Views)** omogućuju podskupove konceptualnog nivoa.
- **Konceptualni nivo (schema)** definiše logičku strukturu **Koji** podaci i **Odnose** između podataka
- **Fizička schema** opisuje **Kako** su podaci memorisani
- (naziva se **ANSI/SPARC** arhitektura)

Korisnici





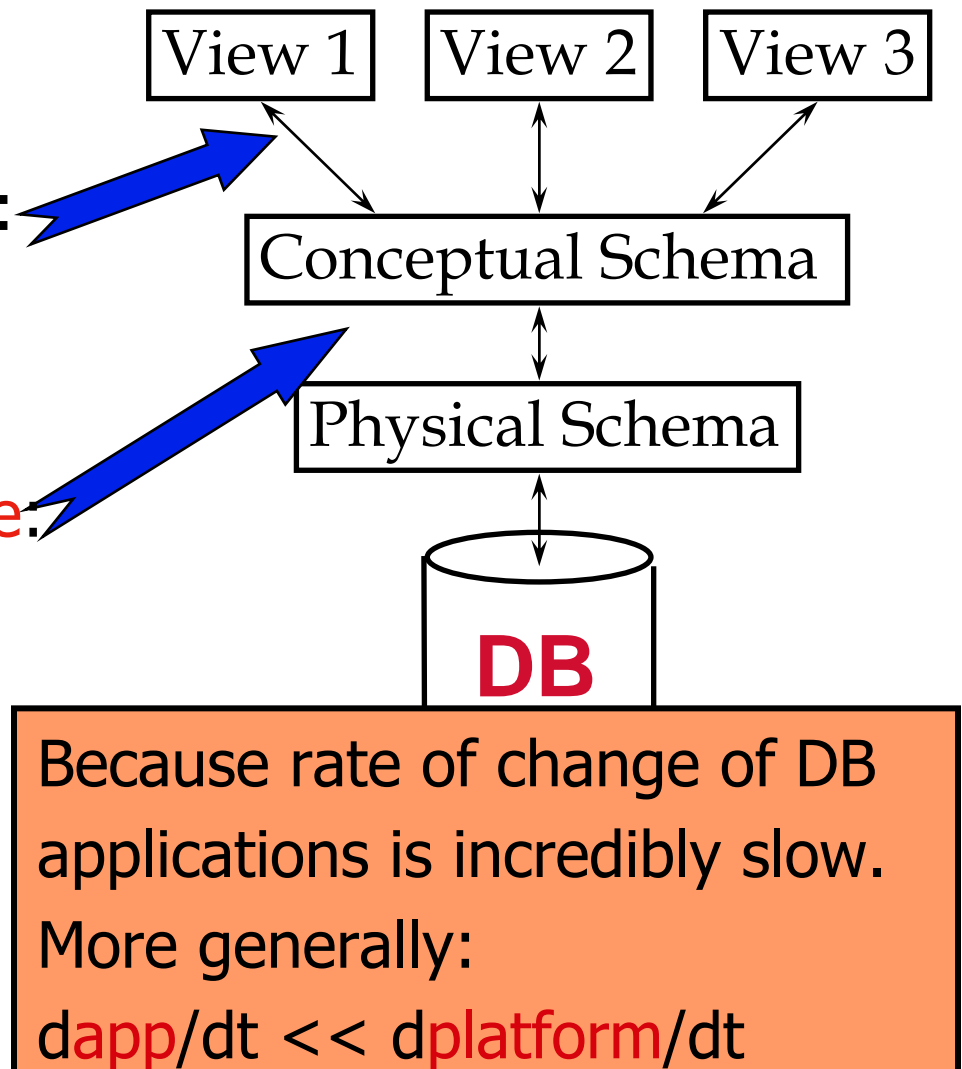
Ortogonalna dimenzija: Šema i Instanca

- Similar to **types** and **variables** in programming languages
- **Schema** – the logical structure of the database
 - Example: The database consists of information about a set of customers and accounts and the relationship between them)
 - Analogous to type information of a variable in a program
 - **Physical schema**: database design at the physical level
 - **Logical schema**: database design at the logical level
- **Instance** – the actual content of the database at a particular point in time
 - Analogous to the value of a variable



Nezavisnost Podataka (Data Independence)

- Applications insulated from how data is structured and stored.
- **Logical data independence:** Protection from changes in logical structure of data.
- **Physical data independence:** Protection from changes in physical structure of data.
- **Q: Why are these particularly important for DBMS?**





Zašto?

- A. Database systems are at the core of CS
- B. They are incredibly important to society
- C. The topic is intellectually rich
- ~~D. It isn't that much work~~
- E. Looks good on your resume

Let's spend a little time on each of these



Zašto?

A. Database systems are the core of CS

- Shift from computation to information
 - True in corporate computing for years
 - Web, p2p made this clear for personal computing
 - Increasingly true of scientific computing
- Need for DB technology has exploded in the last years
 - **Corporate**: retail swipe/clickstreams, “customer relationship mgmt”, “supply chain mgmt”, “data warehouses”, etc.
 - **Web**: not just “documents”. Search engines, e-commerce, blogs, wikis, other “web services”.
 - **Scientific**: digital libraries, genomics, satellite imagery, physical sensors, simulation data
 - **Personal**: Music, photo, & video libraries. Email archives. File contents (“desktop search”).



Zašto?

B. DBs are incredibly important to society

- “Knowledge is power.” --
Sir Francis Bacon
- “With great power comes
great responsibility.” --
SpiderMan’s Uncle Ben

Policy-makers should understand technological possibilities.
Informed Technologists needed in public discourse on usage.

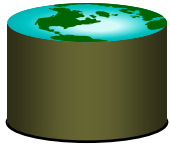


Zašto?

C. The topic is intellectually rich.

- representing information
 - data modeling
- languages and systems for querying data
 - complex queries & query semantics*
 - over massive data sets
- concurrency control for data manipulation
 - controlling concurrent access
 - ensuring transactional semantics
- reliable data storage
 - maintain data semantics even if you pull the plug

* semantics: the meaning or relationship of meanings of a sign or set of signs



Zašto?

~~D. It isn't that much work.~~

- Bad news: It is a lot of work.



Zašto?

E. Looks good on my resume.

- Yes, but why? This is not a course for:
 - Oracle administrators
 - IBM DB2 engine developers
 - Though it's useful for both!
- It is a course for well-educated computer scientists
 - Database system concepts and techniques increasingly used “outside the box”
 - A rich understanding of these issues is a basic and (un?)fortunately unusual skill.



Opis Podataka: Model Podataka

- **Model Podataka** je skup konceptualnih alata za opis:
 - Podataka (Data)
 - Odnosa između podataka (Data relationships)
 - Značenja podataka (Data semantics)
 - Ograničenja (Data constraints)
- **Šema** je opis nekog skupa podataka, korišćenjem datog modela podataka.



Različiti modeli podataka

- Relational model
- Entity-Relationship data model (mainly for database design)
- Object-based data models (Object-oriented and Object-relational)
- Semistructured data model (XML)
- Other older models:
 - Network model
 - Hierarchical model



Primer: Baza Podataka Fakulteta

- Data Model: Relations
- Conceptual schema:
 - **Students**(sid: string, name: string, login: string, age: integer, gpa:real)
 - **Courses**(cid: string, cname:string, credits:integer)
 - **Enrolled**(sid:string, cid:string, grade:string)
- Physical schema:
 - Relations stored as unordered files.
 - Index on first column of Students.
- External Schema (View):
 - **Course_info**(cid:string,enrollment:integer)



Data Manipulation Language (DML)

- Language for accessing and manipulating the data organized by the appropriate data model
 - DML also known as Query Language
- Two classes of languages
 - **Procedural** – user specifies what data is required and how to get those data
 - **Declarative (nonprocedural)** – user specifies what data is required without specifying how to get those data
- SQL is the most widely used Query Language



Data Definition Language (DDL)

- Specification notation for defining the database schema

Example: **create table** *account* (
 account-number **char**(10),
 balance **integer**)

- DDL compiler generates a set of tables stored in a ***Data Dictionary***
- Data dictionary contains metadata (i.e., data about data)
 - **Database schema**
 - **Data *Storage and Definition* language**
 - Specifies the storage structure and access methods used
 - **Integrity constraints**
 - Domain constraints
 - Referential integrity (**references** constraint in SQL)
 - Assertions
 - **Authorization**



SQL

- **SQL**: widely used non-procedural language
 - Example: Find the name of the customer with customer-id 192-83-7465

```
select  customer.customer_name
from    customer
where   customer.customer_id = '192-83-7465'
```
 - Example: Find the balances of all accounts held by the customer with customer-id 192-83-7465

```
select  account.balance
from    depositor, account
where   depositor.customer_id = '192-83-7465' and
        depositor.account_number = account.account_number
```
- Application programs generally access dbase through one of
 - Language extensions to allow embedded SQL
 - Application program interface (e.g., ODBC/JDBC) which allow SQL queries to be sent to a database



Database Architecture

The architecture of a database systems is greatly influenced by the underlying computer system on which the database is running:

- Centralized
- Client-server
- Parallel (multi-processor)
- Distributed



Database Users

Users are differentiated by the way they expect to interact with the system

- **Application programmers** – interact with system through DML calls
- **Sophisticated users** – form requests in a database query language
- **Specialized users** – write specialized database applications that do not fit into the traditional data processing framework
- **Naïve users** – invoke one of the permanent application programs that have been written previously
 - Examples, people accessing database over the web, bank tellers



Database Administrator

- Coordinates all the activities of the database system; the database administrator has a good understanding of the enterprise's information resources and needs.
- Database administrator's duties include:
 - Schema definition
 - Storage structure and access method definition
 - Schema and physical organization modification
 - Granting user authority to access the database
 - Specifying integrity constraints
 - Acting as liaison with users
 - Monitoring performance and responding to changes in requirements



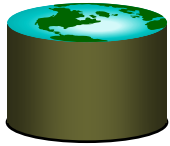
Storage Management

- **Storage manager** is a program module that provides the interface between the low-level data stored in the database and the application programs and queries submitted to the system.
- The SM is responsible to the following tasks:
 - Interaction with the file manager
 - Efficient storing, retrieving and updating of data
- Issues:
 - Storage access
 - File organization
 - Indexing and hashing

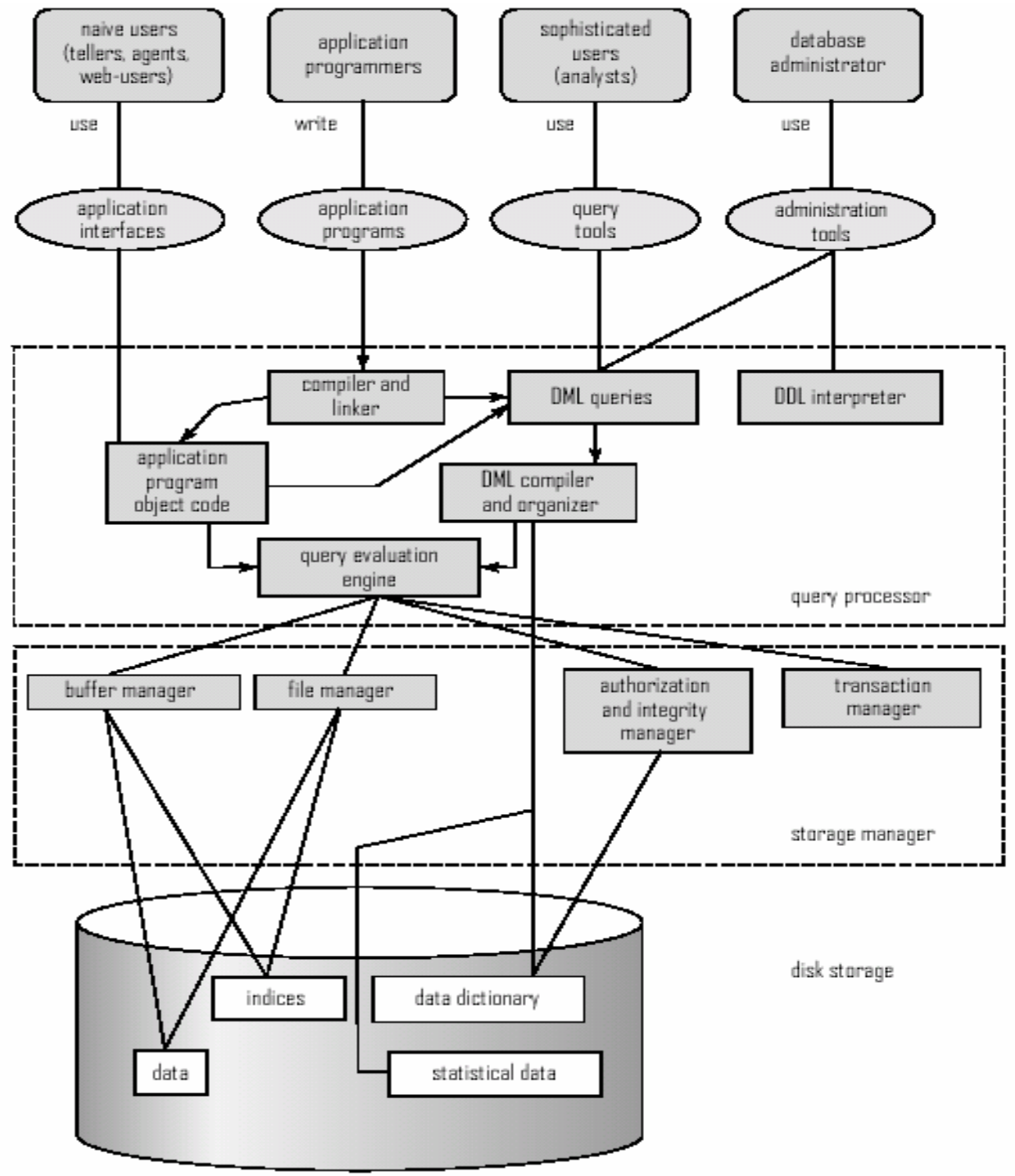


Transaction Management

- A **transaction** is a collection of operations that performs a single logical function in a database application
- **Transaction-management component** ensures that the database remains in a consistent (correct) state despite system failures (e.g., power failures and operating system crashes) and transaction failures.
- **Concurrency-control manager** controls the interaction among the concurrent transactions, to ensure the consistency of the database.



Overall System Structure





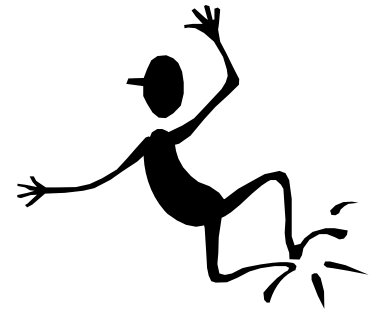
Current Commercial Outlook

- A major part of the software industry:
 - Oracle, IBM, Microsoft
 - also Sybase, Informix (now IBM), Teradata
 - smaller players: java-based dbms, devices, OO, ...
- Well-known benchmarks (esp. TPC)
- Lots of related industries
 - data warehouse, document management, storage, backup, reporting, business intelligence, ERP, CRM, app integration
- Traditional Relational DBMS products dominant and evolving
 - adapting for extensibility (user-defined types), native XML support.
 - Microsoft merging file system/DB for next OS release (??)
- Open Source coming on strong
 - MySQL, PostgreSQL, BerkeleyDB
- And of course, the other “database” technologies
 - Search engines, P2P, etc.



DBMSs make these folks happy ...

- DBMS vendors, programmers
 - Oracle, IBM, MS, Sybase, NCR, ...
- End users in many fields
 - Business, education, science, ...
- DB application programmers
 - Build enterprise applications on top of DBMSs
 - Build web services that run off DBMSs
- Database administrators (DBAs)
 - Design logical/physical schemas
 - Handle security and authorization
 - Data availability, crash recovery
 - Database tuning as needs evolve



...must understand how a DBMS works



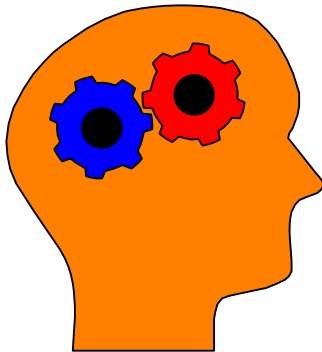
Summary (part 1)

- DBMS used to maintain, query large datasets.
 - can manipulate data and exploit *semantics*
- Most systems over “databases” use related technologies
- Other benefits of DBMSs include:
 - recovery from system crashes,
 - concurrent access,
 - quick application development,
 - data integrity and security.
- Levels of abstraction provide data independence
 - Key when $d_{app}/dt \ll d_{platform}/dt$
- In this course we will explore:
 - 1) How to be a sophisticated user of database technologies
 - Relational
 - 2) What goes on inside a DBMS
 - And related systems



Summary, cont.

- DBAs, DB developers the bedrock of the information economy



- DBMS R&D represents a broad, fundamental branch of the science of computation