

Introduction to Database Management Systems

Daniel POP

Getting around

- Google Classroom: [eryvc3m](#)
- Attendance (physical presence)
 - Lecture: 7
 - Laboratories: 7 (Informatica + Informatica Aplicata), 3 (Matematica Informatica)
 - Otherwise, [recontractare](#)
- Evaluation
 - Lecture: 30% chestionare la curs (nu mai pot fi sustinute in sesiune!), 70% test final in sesiunile A1/B1/C
 - Laborator:
 - test intermediar comanda SELECT
 - proiect individual
 - chestionare / alte mijloace pentru “bonusarea” activitatii si implicarii la laborator

Course Outline

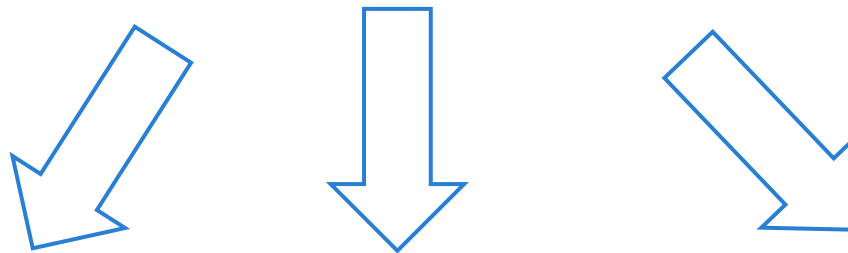
1. Introduction to database approach
2. The database environment
3. Introduction to The Relational Model
4. Views
5. Transactions
6. SQL Constraints.
7. Relational Database Design. Theory and practice
8. An Introduction to Database Performance. Indexing
9. JSON Support in Relational Database Management Systems
10. NoSQL Databases

Week 1

Agenda

1. Why a database approach?
2. Short history of DBMSs
3. The database approach. Key concepts
 1. The database
 2. The Database Management Systems (DBMS)
 3. The application programs
4. Database approach environment
5. Data models
6. Advantages and shortcomings of DBMSs

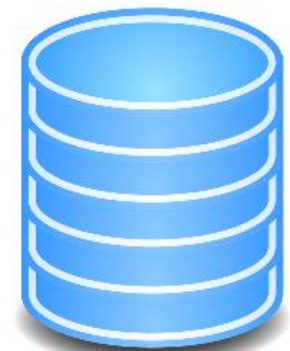
Data storage on different media



Hard-copies on paper

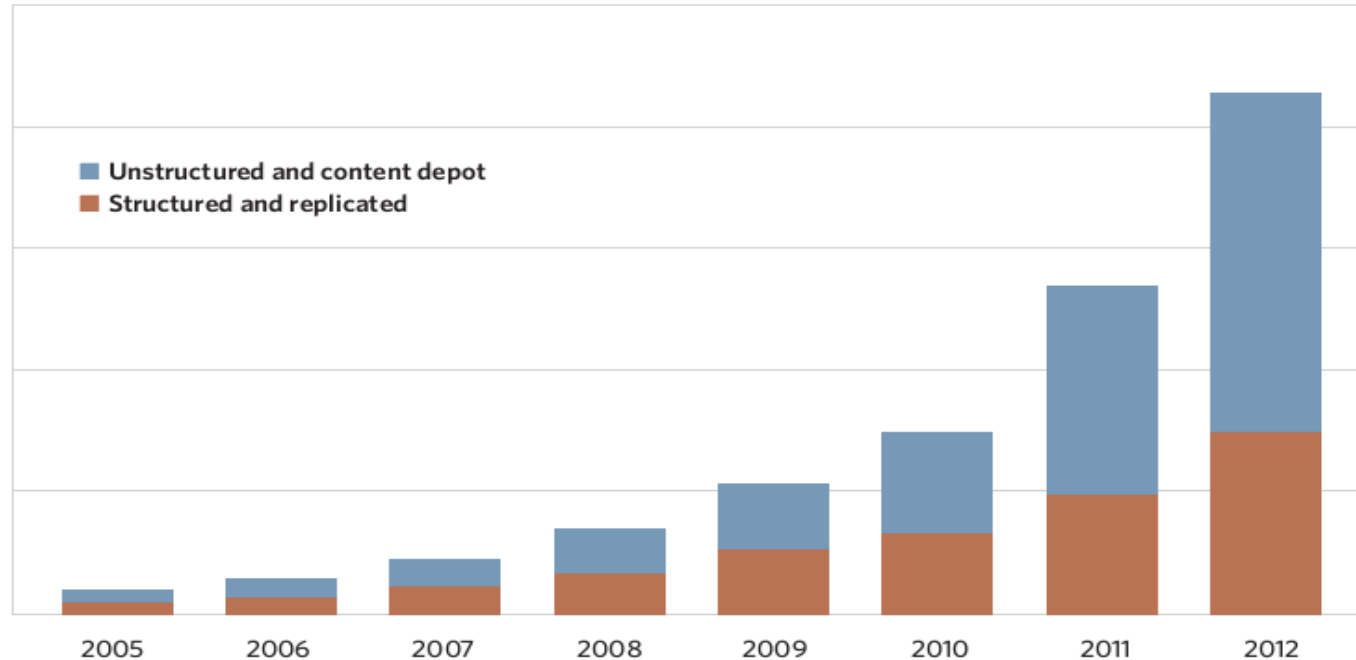


Electronically on files



Database

Massive data



SOURCE: IDC DIGITAL UNIVERSE 2009: WHITE PAPER, SPONSORED BY EMC, 2009.

90% of the data in the world today has been created in the last two years alone.

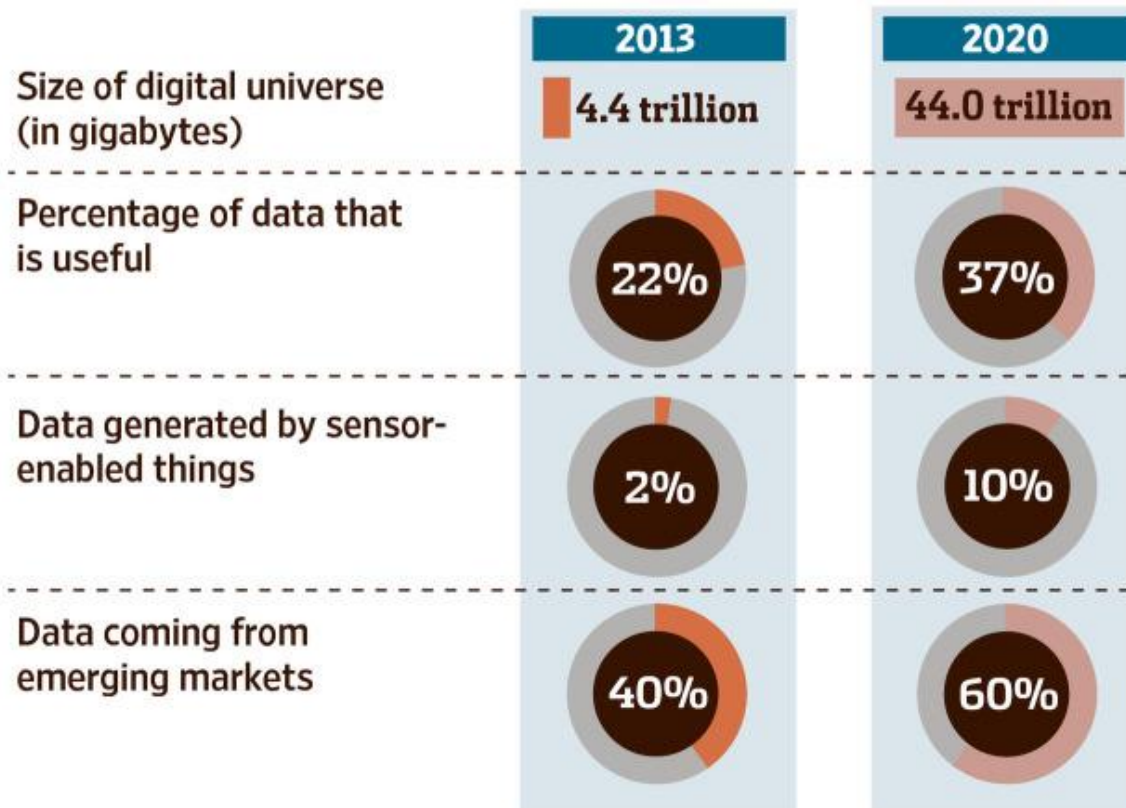
Source: <http://www-01.ibm.com/software/data/bigdata/>

- 12TB of tweets / day
- working with less than 0.001% of the sensor stream data, the data flow from all four LHC experiments represents 25 PB annual rate before replication

...and continues to grow

Data Explosion

The amount of data created and copied annually—known as the digital universe—is projected to expand rapidly this decade, representing an opportunity and challenge for businesses.



Source: IDC, EMC Digital Universe Study

The Wall Street Journal

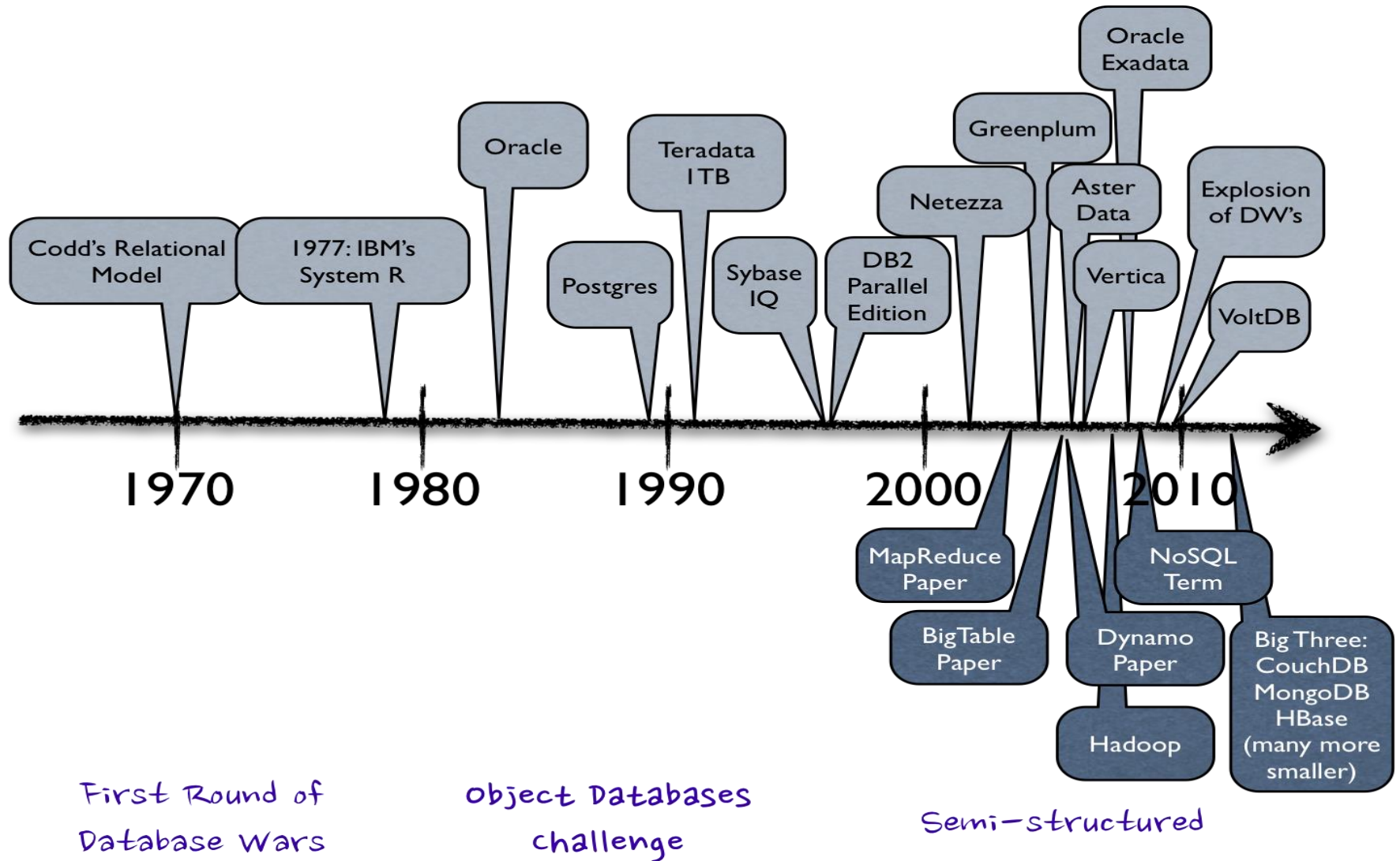
File-based systems

- File-based system = collection of application programs that perform services for end-users such as the production of reports.
- Each program defines and manages its own data.
- A file is a collection of records, containing logically related data. A record contains one or more fields
- Limitations of file-based storage
 - Separation and isolation of data
 - Duplication of data
 - Data dependence
 - Incompatibility of files
 - Fixed queries/proliferation of application programs
 - Inability to generate timely reports

The born of DBMSs

- The DBMS has its roots in the 1960s Apollo moon-landing project - the GUAM (Generalized Update Access Method) - hierarchical approach
- IBM joins in and the IMS (Information Management System) is born
- Mid 1960s GE created IDS (Integrated Data Store) - network DBMS
- 1960-1970 CODASYL (Conference on Data System Languages), DBTG (Data Base Task Group) which define the DDL and the DML
- 1970 Codd introduces the relational model
- 1976 Chen introduces the ER model
- Late 1970s SQL is developed
- Today we have many RDBMSs (DB2 from IBM, Oracle from Oracle, SQL Server from Microsoft etc...)

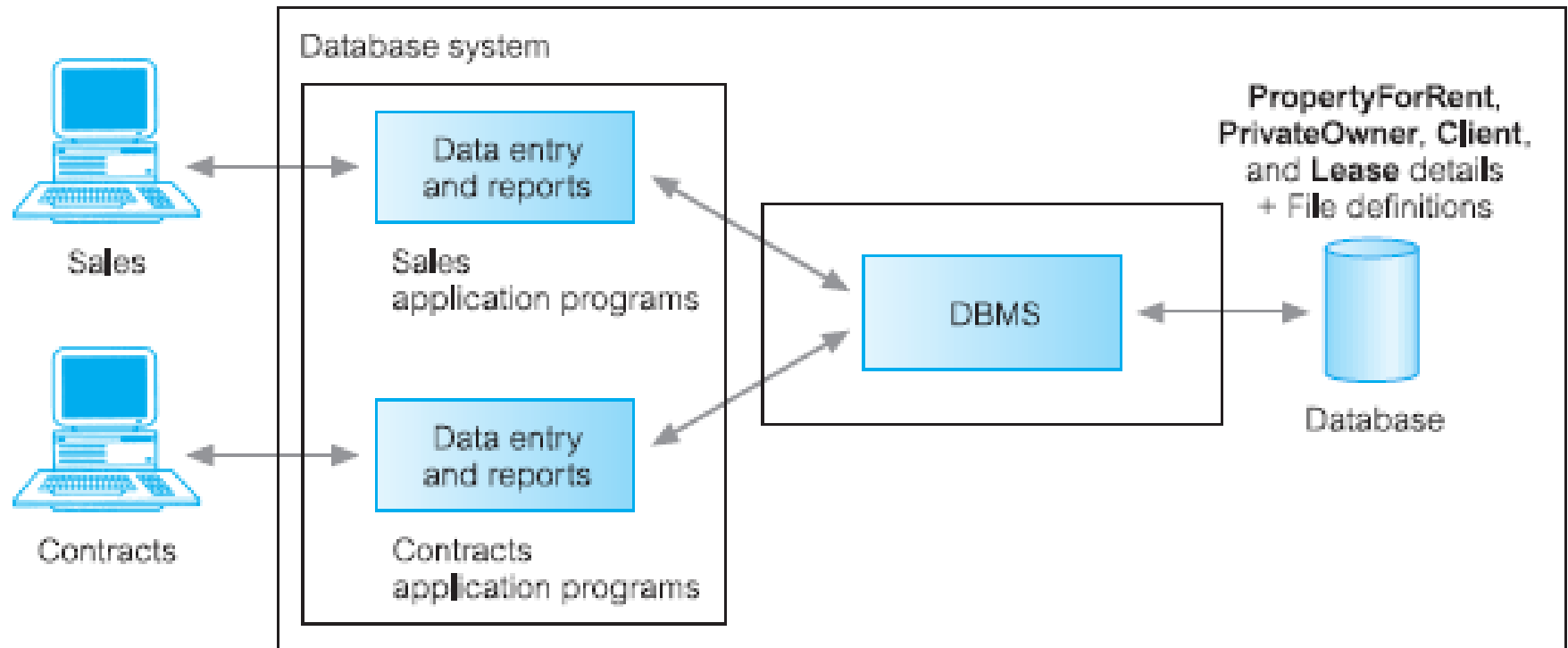
Evolution of DBMS



The Database Approach

- The Database
- The Database Management System
- (Database) Application Programs

The Database Approach



PropertyForRent (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)

PrivateOwner (ownerNo, fName, lName, address, telNo)

Client (clientNo, fName, lName, address, telNo, prefType, maxRent)

Lease (leaseNo, propertyNo, clientNo, paymentMethod, deposit, paid, rentStart, rentFinish)

The Database

- Database = a shared collection of logically related data (and a description of this data), designed to meet the information needs of an organization
- Database = self-describing collection of integrated records
 - Schema = description of data = system catalog, metadata
 - Data model (set of records, XML documents, graph, collection of objects etc.)

Data Models

- **Data Model = integrated collection of concepts for describing data, relationships between data, and constraints on the data in an organization.**
- The purpose of a data model is to represent data and to make the data understandable
- Types of data models
 - External data models
 - Conceptual data models
 - Internal data models

Data Models

- Main roles of data models
 - Communicate the semantic of data
 - Discover the semantic of data
- Characteristics of logical (external/conceptual) data models
 - Graphical diagram
 - Explicit representation of semantic
 - Appropriate level of detail
 - DBMS independent
 - Tool support

Internal data models

- Relational model
- Object-oriented model
- Graph model
- Key-value model
- Columnar model
- Document model

Database Management System

DBMS is a software system that enables users to define, create and maintain the database and provides controlled access to this database.

It provides efficient, reliable, convenient, and safe multi-user storage of and access to massive amounts of persistent data.



Database Management System

- Persistent - Outlive the programs that create/access the data
- Safe - hardware/software failures, malicious users
- Multi-user - concurrently access to data (concurrency control)
- Convenient
 - Physical Data Independence; huge difference between physical representation of data on disk and the logical way of seeing and working with;
 - High level, declarative (what, not how) query languages (e.g. SQL)
- Efficient - thousands of operations (query/update) per second
- Reliable - 99.99999 % uptime

Database Management System

- Allows users to define the database - Data Definition Language (DDL)
- Allows users to insert, update, delete and retrieve data from the database - Data Manipulation Language (DML), e.g. SQL
- Controlled access to the database
 - Security system
 - Integrity system
 - Concurrency control system
 - Recovery control system
 - User-accessible catalog

Database Languages

- Data Definition Language
 - Used to specify the database schema
- Data Manipulation Language
 - Used to update the database (insert, update, delete)
- Data Query Language
 - Used to extract (read) data needed at a moment

Data Definition Language (DDL)

- A descriptive language that allows the DBA or user to describe and name the entities required for the application, together with any associated integrity and security constraints
- The result of the compilation of DDL statements is the system catalog (data dictionary) which integrates metadata
- System catalog (= data dictionary = data directory) may, or may not, be accessible by database users
- DDL used to specify external & conceptual schemas

Data Manipulation Language (DML)

- A language that provides a set of operations to support the basic data manipulation operations on the data held in the database
- Operations include insertion, modification or deletion
- Procedural DML - a language that allows the user to tell the system what data is needed and exactly **how** to retrieve the data (network and hierarchical DMLs are typically procedural)
- Declarative DML - a language that allows the user to state **what** data is needed rather than how it is to be retrieved

Data Query Language (DQL)

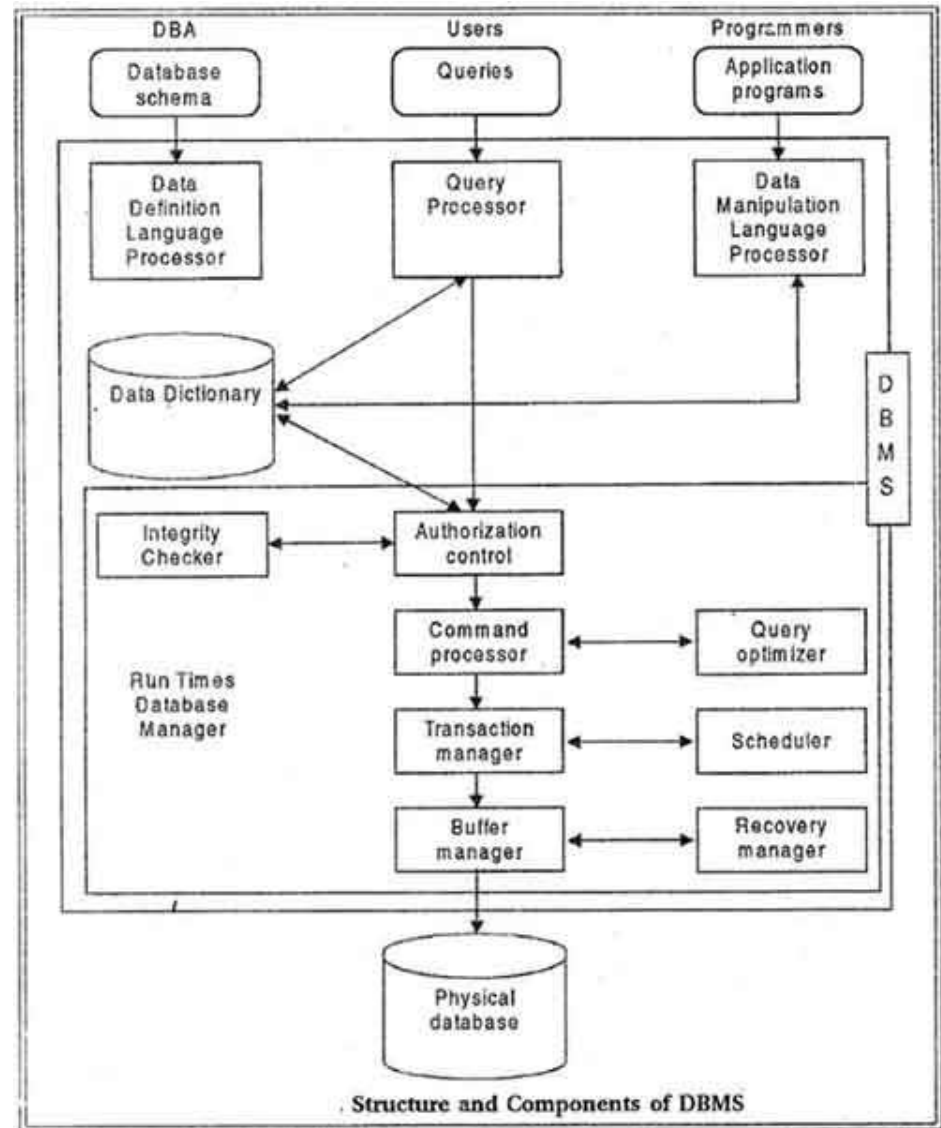
- A language that provides a set of operations to support the basic data extraction on the data held in the database
- Operations include querying the database
- Procedural DQL vs. Declarative DQL

4GLs

- Non procedural
- Presentation languages (query languages, report generators)
- Specialty languages (spreadsheets)
- Application / forms / graphics generators
- Example: SQL, QBR (Query By Example)

Components of a DBMS

- Query processor
- DML preprocessor
- DDL compiler
- Data dictionary
- Run-time database manager

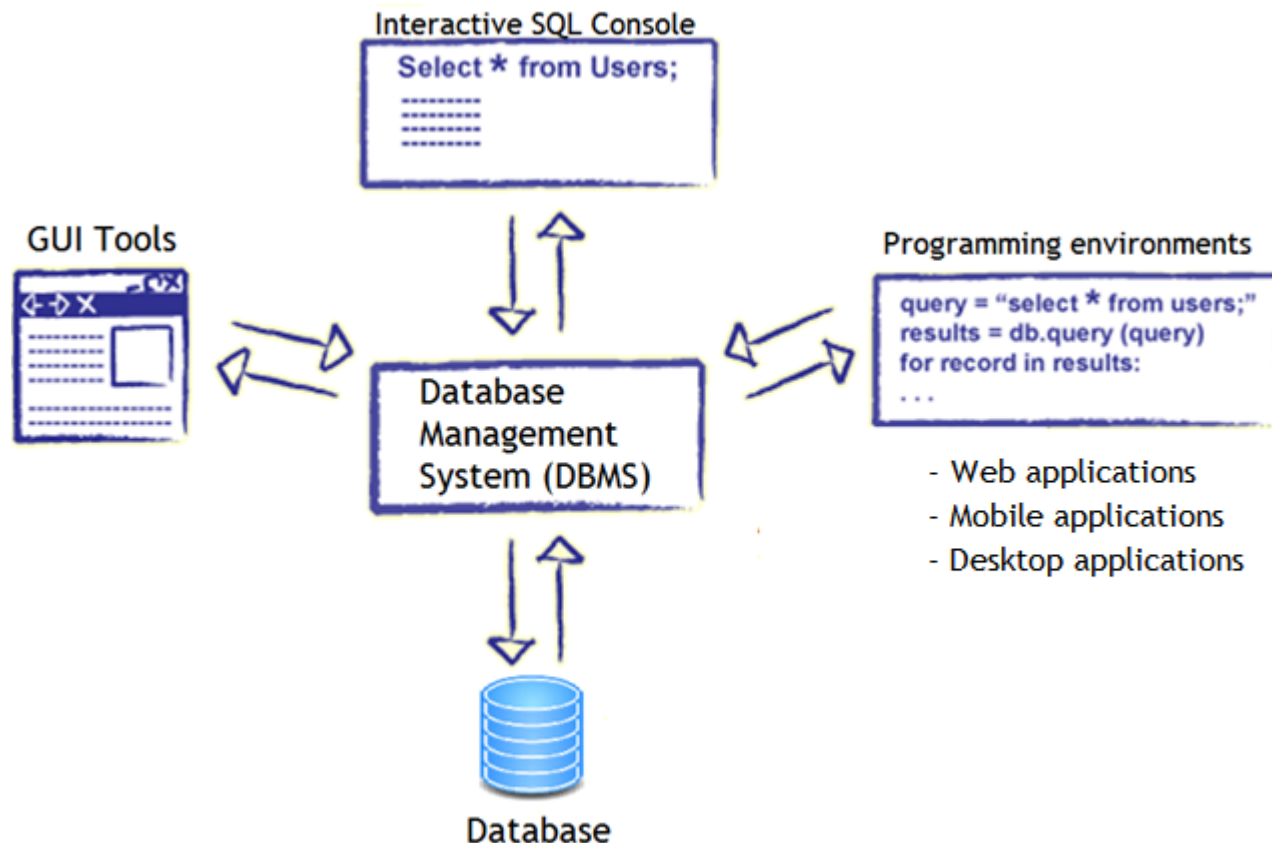


Show me a DBMS

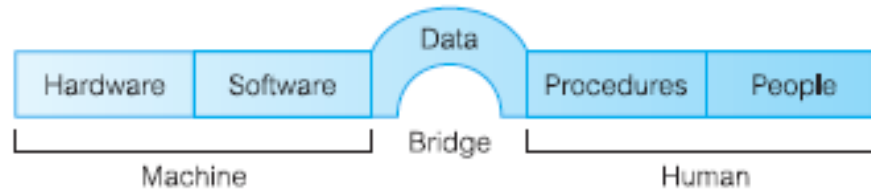
Services (Local)					
Select an item to view its description.					
Name	Description	Status	Startup Type	Log On As	
Sensor Service	A service for sensors that manages different sensors' functionality. Manages S...		Manual (Trigger Start)	Local System	
Server	Supports file, print, and named-pipe sharing over the network for this compu...	Running	Automatic (Trigger Start)	Local System	
Shared PC Account Manager	Manages profiles and accounts on a SharedPC configured device	Disabled	Automatic	Local System	
Shell Hardware Detection	Provides notifications for AutoPlay hardware events.	Running	Automatic	Local System	
Smart Card	Manages access to smart cards read by this computer. If this service is stoppe...	Running	Manual (Trigger Start)	Local Service	
Smart Card Device Enumeration Service	Creates software device nodes for all smart card readers accessible to a given ...		Manual (Trigger Start)	Local System	
Smart Card Removal Policy	Allows the system to be configured to lock the user desktop upon smart card ...		Manual	Local System	
SNMP Trap	Receives trap messages generated by local or remote Simple Network Manag...		Manual	Local Service	
Software Protection	Enables the download, installation and enforcement of digital licenses for Wi...	Running	Automatic (Delayed Start, Trigg...	Network Se...	
Spatial Data Service	This service is used for Spatial Perception scenarios		Manual	Local Service	
Spot Verifier	Verifies potential file system corruptions.		Manual (Trigger Start)	Local System	
SQL Full-text Filter Daemon Launcher (MSSQLSERVER)	Service to launch full-text filter daemon process which will perform document...	Running	Manual	NT Service\...	
SQL Server (MSSQLSERVER)	Provides storage, processing and controlled access of data, and rapid transacti...	Running	Automatic	Local System	
SQL Server Agent (MSSQLSERVER)	Executes jobs, monitors SQL Server, fires alerts, and allows automation of som...		Manual	NT Service\...	
SQL Server Analysis Services (MSSQLSERVER)	Supplies online analytical processing (OLAP) and data mining functionality for...		Disabled	NT Service\...	
SQL Server Analysis Services CEIP (MSSQLSERVER)	CEIP service for Sql Server Analysis Services	Running	Automatic	NT Service\...	
SQL Server Browser	Provides SQL Server connection information to client computers.		Disabled	Local Service	
SQL Server CEIP service (MSSQLSERVER)	CEIP service for Sql server	Running	Automatic	NT Service\...	
SQL Server Launchpad (MSSQLSERVER)	Service to launch Advanced Analytics Extensions Launchpad process that ena...	Running	Automatic	NT Service\...	
SQL Server PolyBase Data Movement (MSSQLSERVER)	Manages communication and data transfer between SQL Server and external ...		Disabled	Network Se...	
SQL Server PolyBase Engine (MSSQLSERVER)	Creates, coordinates and executes the parallel query plan against external dat...		Disabled	Network Se...	
SQL Server VSS Writer	Provides the interface to backup/restore Microsoft SQL server through the Wi...	Running	Automatic	Local System	
SSDP Discovery	Discovers networked devices and services that use the SSDP discovery protoc...	Running	Manual	Local Service	
State Repository Service	Provides required infrastructure support for the application model.	Running	Manual	Local System	
Still Image Acquisition Events	Launches applications associated with still image acquisition events.		Manual	Local System	
Storage Service	Provides enabling services for storage settings and external storage expansion	Running	Automatic (Delayed Start, Trigg...	Local System	
Storage Tiers Management	Optimizes the placement of data in storage tiers on all tiered storage spaces i...		Manual	Local System	
Sync Host_1b33ce	This service synchronizes mail, contacts, calendar and various other user data. ...	Running	Automatic (Delayed Start)	Local System	
SysMain	Maintains and improves system performance over time.	Running	Automatic	Local System	
System Event Notification Service	Monitors system events and notifies subscribers to COM+ Event System of th...	Running	Automatic	Local System	
System Events Broker	Coordinates execution of background work for WinRT application. If this servi...	Running	Automatic (Trigger Start)	Local System	
System Guard Runtime Monitor Broker	Monitors and attests to the integrity of the Windows platform.	Running	Automatic (Delayed Start, Trigg...	Local System	
Task Scheduler	Enables a user to configure and schedule automated tasks on this computer. ...	Running	Automatic	Local System	
TCP/IP NetBIOS Helper	Provides support for the NetBIOS over TCP/IP (NetBT) service and NetBIOS na...	Running	Manual (Trigger Start)	Local Service	

(Database) Application Programs

Application Program = a computer program that interacts with the database by issuing an appropriate request (typically an SQL statement) to the DBMS.

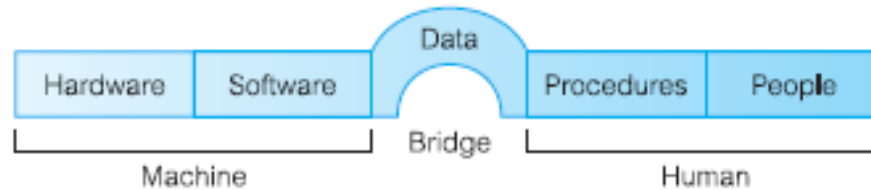


Database Approach Environment



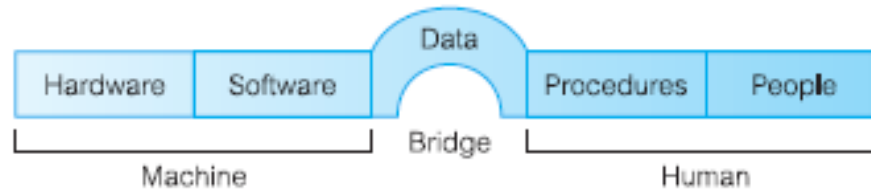
- Hardware
 - Single server
 - Distributed architecture
- Software (OS, DBMS, application programs)
 - Apps are usually written in 3GL (e.g. C++, Java, Visual Basic, PHP)
 - The DBMS may have 4GL for query languages, report generators, form generators, graphics generators, application generators (e.g. MS Access, SQL)

Database Approach Environment



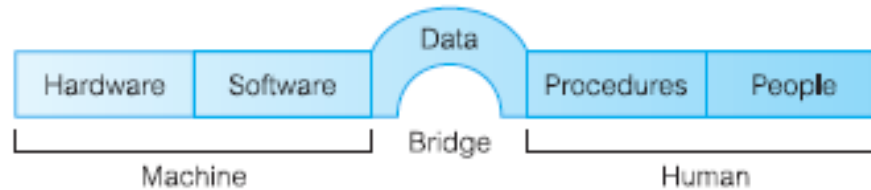
- Data (schema = structure of the database, tables, attributes)
 - Data models
 - Names, types, and sizes of data items
 - Names of relationships
 - Integrity constraints
 - Names of authorized users
 - Indexes, storage structures

Database Approach Environment



- Procedures
 - Instructions and rules that govern the design and use of the database
 - log on to the DBMS;
 - use a particular DBMS facility or application program;
 - start and stop the DBMS;
 - make backup copies of the database;
 - handle hardware or software failures
 - coding standards, guidelines
 - monitoring + notifications
 - migration

Database Approach Environment



- People (Roles in the Database Environment)
 - Data administrator (DA): planning, development and maintenance of standards, policies and procedures + conceptual/logical design
 - Database administrator (DBA): physical design and implementation, security, integrity, maintenance of the OS, ensuring satisfactory performance for apps and users
 - Database designer: logical/conceptual database designer - business rules, physical database designer
 - Application programmers
 - End-users
 - Naïve users
 - Sophisticated users

Advantages of DBMSs

- Flexibility: more information from the same amount of data
- Control of data redundancy
- Data consistency
- Sharing of data
- Improved data integrity
- Improved security
- Enforcement of standards
- Economy of scale
- Balance of conflicting requirements
- Improved data accessibility and responsiveness
- Increased productivity
- Improved maintenance
- Increased concurrency
- Improved backup and recovery services

Shortcomings of DBMSs

- Complexity and cost (HW + DBMS costs + cost of conversion to DB approach)
- Size
- Performance
- Higher impact of a failure

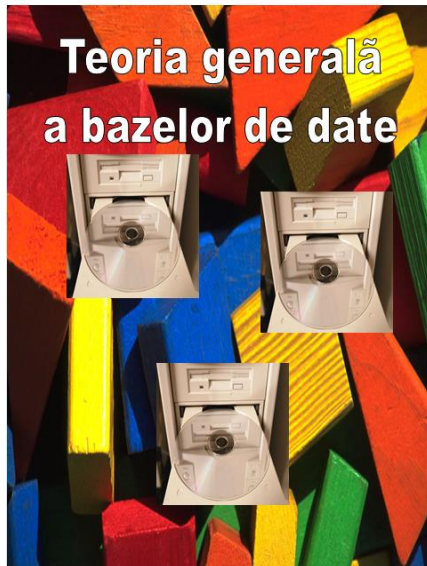
Common uses of DBMS

- Web applications
- Super/Hypermarket
- Credit card management
- Library
- Insurance
- Manufacturing
- Financial/accounting
- Social media

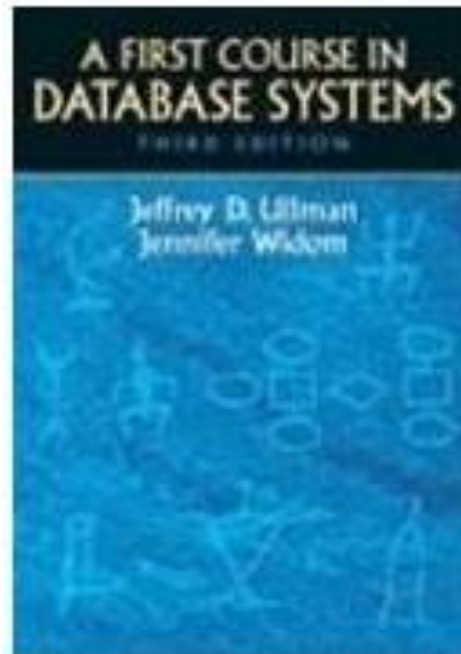
Bibliography (recommended)

JOAN DESPI
GHEORGHE PETROV

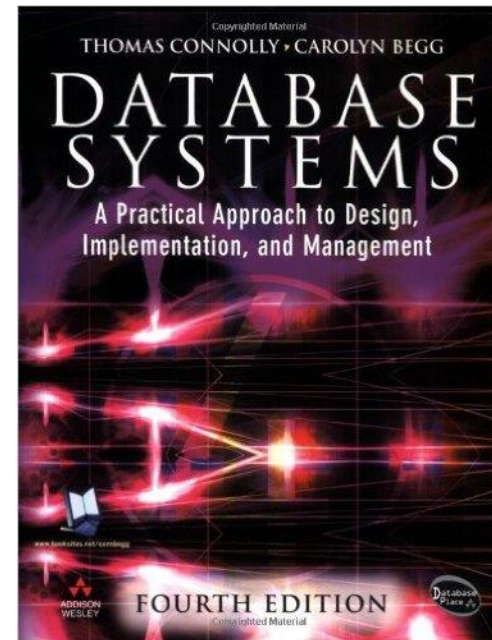
REISZ ROBERT
AUREL STEPAN



Teoria generala a bazelor de date,
I. Despi, G.
Petrov, R. Reisz,
A. Stepan,
Mirton, 2000
Cap 1 & 2

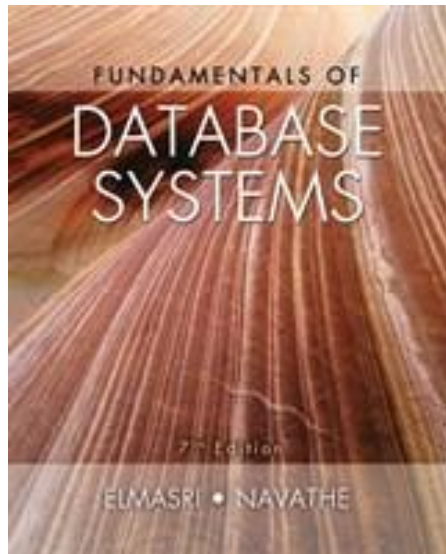


*A First Course in
Database Systems
(3rd edition) by
Jeffrey Ullman and
Jennifer Widom,
Prentice Hall, 2007*
Chapter 1

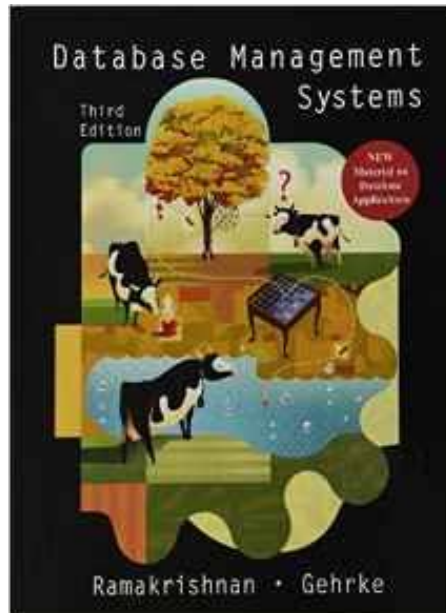


*Database Systems - A
Practical Approach to
Design, Implementation,
and Management (4th
edition) by Thomas
Connolly and Carolyn
Begg, Addison-Wesley,
2004*
Chapter 1 & 2

Bibliography (optional)



Fundamentals of Database Systems (6th edition) by Ramez Elmasri and Shamkant Navathe, Addison-Wesley, 2010



Database Management Systems (3rd edition) by Raghu Ramakrishnan and Johannes Gehrke, McGraw-Hill, 2002