Database Systems

Haitham A. El-Ghareeb

July 5, 2019

Faculty of Computers and Information Sciences
Mansoura University
Egypt
helghareeb@mans.edu.eg

Contacts

- https://youtube.com/helghareeb
- https://www.github.com/helghareeb
- http://eg.linkedin.com/in/helghareeb
- h.elghareeb@yahoo.com
- https://www.haitham.ws

Objectives

Objectives

After completing this chapter, you will be able to:

1. Define the difference between data and information
2. Describe what a database is, the various types of databases, and why they are valuable assets for decision making
3. Explain the importance of database design
4. See how modern databases evolved from file systems
5. Understand flaws in file system data management
6. Outline the main components of the database system
7. Describe the main functions of a database management system (DBMS)

Preview

Preview

1. Organizations use data to keep track of their day-to-day operations
2. Such data is used to generate information
3. Databases evolved from the need to manage large amounts of data in an organized and efficient manner
4. Computer file systems were used to organize data
5. Though outmoded, file systems are the source of serious data management limitations

Why Databases?

Data is

1. Ubiquitous [1] (abundant [2], global, everywhere)
2. Pervasive [3] (prevalent [4], persistent)
3. Essential for organizations to survive and prosper

[1] present, appearing, or found everywhere
[2] existing or available in large quantities
[3] spreading widely throughout an area or a group of people.
[4] widespread in a particular area or at a particular time

Business Information Systems

1. It can be argued that the ultimate purpose of all business information systems is to help businesses use information as an organizational resource.
2. At the heart of all these systems are the [. . .] of data
   - collection
   - storage
   - aggregation [5]
   - manipulation
   - dissemination [6]
   - management

[5] the formation of a number of things into a cluster

[6] the action or fact of spreading something, especially information, widely

Why Databases?

Examples

Telecommunications Companies

1. Keep data on trillions of phone calls
2. new data being added to the system at speeds up to 70,000 calls per second
3. They have to
    - store
    - manage
    - find facts in that data quickly

Google

1. No official information!
2. Estimates
    - 91 million searches per day
    - data: several terabytes
    - results are available almost instantly

Why Databases?

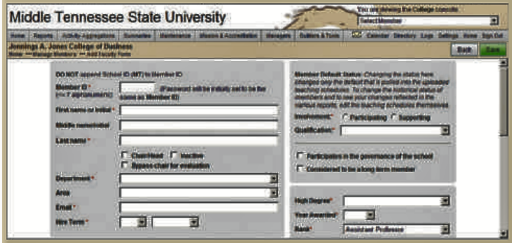
Data vs. Information

Data

1. raw facts
2. raw: facts have not yet been processed to reveal their meaning
3. raw state: the format in which data was collected
4. University Information System: List of Doctors, Students, Courses, etc.
5. Raw data must be properly *formatted* for storage, processing, and presentation
   - Dates: Stored in Julian calendar format, displayed in many formats
   - Yes / No: Stored in 0/1 value format
6. More complex formatting is required when working with complex data types, such as sounds, videos, or images

Further Processing of Data

a) Data entry screen



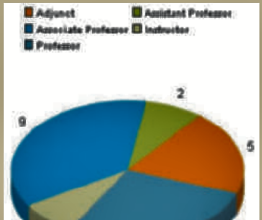
b) Raw data

Id	LastName	MidName	FirstName	DeptCode	Office	Email	Rank	HireYear	Degree
1	Washington	A.	George	MGMT	N136	gwashington@mtsu.edu	Professor	2001	Ph.D.
2	Adams		John	FIN	N213	jadams@mtsu.edu	Professor	1984	Ph.D.
3	Jefferson	L.	Thomas	ECON		tjefferson@mtsu.edu	Instructor	2002	M.B.A.
4	Madison	D.	James	FIN	N236	jmadison@mtsu.edu	Associate Professor	1984	Ph.D.
5	Monroe	R.	James	ACCT	N411	jmonroe@mtsu.edu	Assistant Professor	1995	Ph.D.
6	Adams	Q.	John	ACCT	N410	jqadams@mtsu.edu	Associate Professor	1989	Ph.D.
7	Jackson	C.	Andrew	ECON	N367	ajackson@mtsu.edu	Associate Professor	1989	Ph.D.
8	Van Buren	T.	Martin	FIN	N386	mvanburen@mtsu.edu	Professor	1998	Ph.D.
9	Harrison	R.	William	MKTG	N130	wharrison@mtsu.edu	Professor	1984	Ph.D.
10	Tyler	M.	John	MGMT		jtyler@mtsu.edu	Assistant Professor	2000	Ed.D.
11	Polk		Cheryl	MKTG	N141	cpolk@mtsu.edu	Associate Professor	2002	Ph.D.
12	Taylor	G.	Zachary	ACCT	N415	ztaylor@mtsu.edu	Associate Professor	1986	Ph.D.
13	Filmore		Millard	JCB	N219	mfilmore@mtsu.edu	Professor	1992	Ph.D.
14	Pierce	A.	Franklin	MKTG	N250	fpierce@mtsu.edu	Instructor	2005	M.B.A.
15	Buchanan	T.	James	MGMT	N146	jbuchanan@mtsu.edu	Associate Professor	1998	D.B.A.
16	Lincoln	W.	Larry	MGMT	N150	llincoln@mtsu.edu	Associate Professor	1998	Ph.D.
17	Johnson		Andrew	SYS	N260	ajohnson@mtsu.edu	Professor	1987	Ph.D.
18	Grant		Katie	MKTG	N120	kgrant@mtsu.edu	Assistant Professor	1999	D.B.A.
19	Rutherford		Hayes	ACCT	N488	hrutherford@mtsu.edu	Professor	1992	Ph.D.
20	Garfield	T.	Denise	ACCT		dgarfield@mtsu.edu	Assistant Professor	2018	Ph.D.
21	Arthur		Emily	ACCT	N413	earthur@mtsu.edu	Associate Professor	2002	J.D.
22	Cleveland	G.	Robert	ACCT	N481	rcleveland@mtsu.edu	Associate Professor	1987	Ph.D.
24	Harrison	X.	Patricia	BULA	N486	pharrison@mtsu.edu	Associate Professor	2001	J.D.
25	McKinley	B.	Priscilla	SYS	N363	pmckinley@mtsu.edu	Adjunct	1994	M.S.
26	Roosevelt	F.	Hilary	MGMT	N184	hroosevelt@mtsu.edu	Associate Professor	2002	Ph.D.
27	Wilson		Laura	BCEN	N448	lwilson@mtsu.edu	Professor	1992	Ph.D.
28	Harding		Warren	MKTG	N114	wharding@mtsu.edu	Professor	1984	Ed.D.
29	Coolidge		Calvin	ECON	N339	ccoolidge@mtsu.edu	Professor	1975	Ph.D.
30	Hoover		Lisa	MGMT		lhoover@mtsu.edu	Adjunct	1978	M.B.A.
31	Truman		Betty	ACCT	N416	btruman@mtsu.edu	Professor	1971	Ed.D.
32	Johnson		Robert	BCEN	N240	rjohnson@mtsu.edu	Professor	2001	Ph.D.

c) Information in summary format

Rank	COUNT	%INFS	TOT/COL	%/COL. TOT.	%/COL. FAC.
Adjunct	5	20.00%	23	21.74%	3.27%
Assistant Professor	2	8.00%	28	7.14%	1.31%
Associate Professor	9	36.00%	37	24.32%	5.88%
Instructor	2	8.00%	18	11.11%	1.31%
Professor	7	28.00%	47	14.89%	4.58%

d) Information in graphical format



Extract Meaningful Data

1. What percentage of the faculty in the Information Systems department are adjuncts [7] ?
2. What is Professor to Students Ratio in Information Systems Department?
3. Graphics enhance ability to quickly extract meaning from data

[7] a person who is another's assistant or subordinate

Information

1. the result of processing raw data to reveal its meaning
2. Data processing can be as simple as organizing data to reveal patterns
3. or as complex as making forecasts or drawing inferences using statistical modeling
4. foundation for decision making
5. Information must be: accurate, relevant [8], and timely [9]
6. To reveal meaning, information requires *context*

[8] closely connected or appropriate to what is being done or considered

[9] appropriate to the current time, period, or circumstances; of contemporary (occurring at the same time) interest

Context Examples

1. Temperature reading of 105 degrees does not mean much
    - Celsius vs. Fahrenheit
    - Machine or body or outside air temperature

Knowledge

1. the body of information and facts about a specific subject
2. impacts familiarity, awareness, and understanding of information as it applies to an environment
3. can be thought of as: new knowledge derived from knowledge

Why Databases?

Definitions

Data vs. Information vs. Knowledge

1. **Data** Raw facts, or facts that have not yet been processed to reveal their meaning to the end user
2. **Information** The result of processing raw data to reveal its meaning. Information consists of transformed data and facilitates decision making
3. **Knowledge** The body of information and facts about a specific subject. Knowledge implies familiarity, awareness, and understanding of information as it applies to an environment. A key characteristic is that new knowledge can be derived from old knowledge

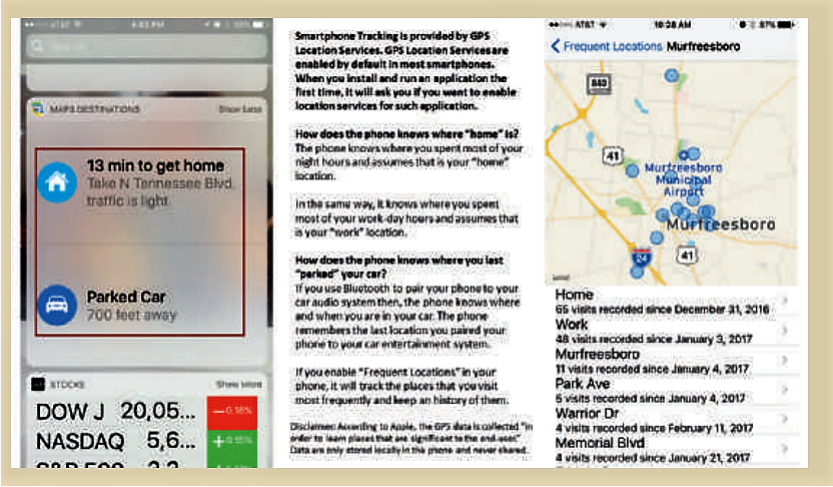
Why Databases?

Summary

Summary

1. Data constitutes the building blocks of information.
2. Information is produced by processing data.
3. Information is used to reveal the meaning of data.
4. Accurate, relevant, and timely information is the key to good decision making.
5. Good decision making is the key to organizational survival in a global environment.

Data vs. Information vs. Knowledge Example



Data Management

1. Discipline that focuses on the proper generation, storage, and retrieval of data

Introducing the Database

Database

1. Shared, integrated computer structure that stores a collection of the following:
   1.1 End-user data: raw facts of interest to the end user
   1.2 Metadata: or data about data, through which the end-user data is integrated and managed

Metadata Examples

1. Describes the data characteristics and the set of relationships that links the data found within the database.
2. Stores information such as:
   2.1 Name of each data element
   2.2 Type of values (numeric, dates, or text) stored on each data element
   2.3 Whether the data element can be left empty
3. Provides information that complements and expands the value and use of the data
4. Presents a more complete picture of the data in the database

Self-Describing Data

1. Given the characteristics of metadata, the database is described as
    - collection of self-describing data

Database Management System

1. DBMS
2. Collection of programs that manages the database structure and control access to the data stored in the database
3. Database: resembles a very well-organized electronic filling cabinet in which powerful software (DBMS) helps manage the cabinet's contents

Introducing the Database

Definitions

Data Management

1. A process that focuses on data collection, storage, and retrieval.
2. Common data management functions include addition, deletion, modification, and listing

Database

1. A shared, integrated computer structure that houses a collection of related data.
2. A database contains two types of data: end-user data (raw facts) and metadata

Metadata

1. Data about data; that is, data about data characteristics and relationships.

DBMS

1. Database Management System
2. The collection of programs that manages the database structure and controls access to the data store in the database.

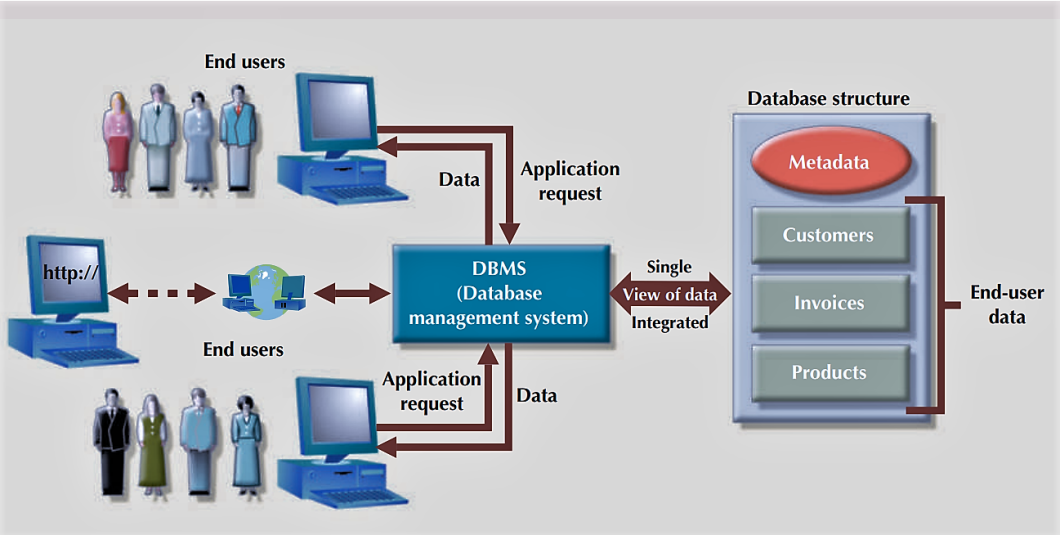
Introducing the Database

DBMS Role

DBMS Role

1. Serves as the intermediary between the user and the database
2. DB structure itself is stored as a collection of files, and the only way to access the data in those files is through the DBMS
3. Presents the end user (application program) with a single, integrated view of the data in the database
4. Receives all application requests and translates them into the complex operations required to fulfill those requests
5. Hides much of the DB's complexity from the application programs and the user

DBMS Role


End users
Data
Application request
Database structure
Metadata
Customers
Invoices
Products
End-user data
http://
DBMS (Database management system)
Single View of data Integrated
End users
Application request
Data

DBMS Advantages

1. enables the data in the database to be *shared* among multiple applications or users
2. *integrates* the many different users' views of the data into a single all-encompassing data repository

DBMS Advantages - 01

1. *Improved data sharing*
2. *Improved data security*
3. *Better data integration*
4. *Minimized data inconsistency* - Data inconsistency: exists when different versions of the same data appear in different places
5. *Improved data access*
   - 5.1 Query: specific request issued to the DBMS for data manipulation - example: read or update data
   - 5.2 Query: simply a Question
   - 5.3 ad hoc query: is a spur-of-the-moment question / on impulse; without planning in advance.
   - 5.4 DBMS sends back an answer - called the Query Result Set

DBMS Advantages - 02

6. *Improved decision making* - Data Quality: comprehensive approach to promoting the accuracy, validity, and timeliness of the data [10]
7. *Increased end-user productivity*
8. Many more!

[10] DBMS does not guarantee data quality, it provides framework to facilitate data quality initiatives

Introducing the Database

Definitions

Data Inconsistency

1. A condition in which different versions of the same data yield different (inconsistent) results

Query

1. A question or task asked by an end user of a database in the form of SQL code.
2. A specific request for data manipulation issued by the end user of the application to the DBMS

ad hoc query

1. A "spur-of-the-moment" question
2. on impulse; without planning in advance.

Query Result Set

1. The collection of data rows returned by a query

Data Quality

1. A comprehensive approach to ensuring the accuracy, validity, and timeliness of data

Types of Databases

Classifications

1. DBMS can be used to build many different types of databases
2. Each database stores a particular collection of data and is used for a specific purpose
3. Different criteria for classification
   3.1 number of users supported
   3.2 where data is located
   3.3 type of data stored
   3.4 intended data usage
   3.5 degree to which the data is structured

Types of Databases

Classification 01: Number of Users

Classification 01: Number of Users

1. Single-User DB
2. Multi-User DB

Single-User DB

1. supports only one user at a time
2. If user A is using the database, users B and C must wait until user A is done
3. Single-User DB that runs on a personal computer is called a *desktop database*

Multi-User Database

1. Supports multiple users at the same time
2. When DB supports a relatively small no. of users (usually fewer than 50) or a specific department within an organization, it is called *Workgroup DB*
3. When DB is used by the entire organization and supports many users (more than 50, usually hundreds) across many departments, the DB is known as *Enterprise DB*

Types of Databases

Classification 02: Location

Classification 02: Location

1. Centralized DB
2. Distributed DB
3. Cloud DB

Centralized DB

1. DB that supports data located at a single site

Distributed DB

1. DB that supports data distributed across several different sites

Cloud DB

1. In Centralized and Distributed DB, infrastructure is owned and maintained by the organization
2. Cloud DB: DB that is created and maintained using cloud data services
3. Examples: Microsoft Azure, Amazon AWS

Types of Databases

Classification 03: Type of Data Stored

Classification 03: Type of Data Stored

1. General Purpose DB
2. Discipline-Specific DB

General Purpose DB

1. contain a wide variety of data used in multiple disciplines
2. Examples:
   2.1 Census [11] DB: contains general demographic [12] data
   2.2 LexisNexis [13] and ProQuest [14] DB: contain newspaper, magazine, and journal articles for a variety of topics

[11] an official count or survey, especially of a population
[12] relating to the structure of populations
[13] `https://www.lexisnexis.com`
[14] `http://www.proquest.com`

Discipline-Specific DB

1. contain data focused on specific subject areas
2. Data is used mainly for academic or research purposes within a small set of disciplines
3. Examples
   3.1 Financial Data: CompuStat or Center for Research in Security Prices (CRSP) [15]
   3.2 Geographic Information Systems (GIS): store geo-spatial and other related data
   3.3 Medical DB: store confidential medical history data

[15] http://www.crsp.com

Types of Databases

Classification 04: Intended Data Usage

Classification 04: Intended Data Usage

1. Most common classification
2. Based on
   2.1 how data will be used
   2.2 and time sensitivity of the information gathered from them
3. Classified into
   3.1 Operational DB
   3.2 Analytical DB

Operational DB

1. Online Transaction Processing (OLTP) or Production DB
2. Transactions such as product or service sales, payments, and supply purchases reflect critical day-to-day operations
3. must be recorded accurately and immediately

Analytical DB

1. focuses primarily on storing historical data and business metrics
2. used exclusively for tactical [16] or strategic [17] decision making
3. Requires extensive "data massaging / data manipulation" to produce information on which to base pricing decisions, sales forecasts, market strategies, and so on
4. allow end user to perform advanced analysis of business data using sophisticated tools
5. Typically comprises two main components
   5.1 Data Warehouse
   5.2 Online Analytical Processing (OLAP)

[16] medium term decisions

[17] long-term goals and a longer term vision

Data Warehouse

1. Specialized DB
2. Stores data in a format optimized for decision support
3. contains historical data obtained from the operational DBs as well as data from other external sources

Online Analytical Processing (OLAP)

1. set of tools that work together
2. provide an advanced data analysis environment for [...] data from the data warehouse
   - retrieving
   - processing
   - modeling
3. Grown into: Business Intelligence (BI)

Business Intelligence

1. Describes a comprehensive approach to capture and process business data
2. with the purpose of generating information
3. to support business decision making

Types of Databases

Classification 05: Degree to which Data is Structured

Classification 05: Degree to which Data is Structured

1. Unstructured Data
2. Structured Data
3. Semi-Structured Data

Unstructured Data

1. Data that exists in its original (raw) state
2. Raw state = in the format in which it was collected
3. Exists in a format that does not lend itself to the processing that yields information

Structured Data

1. the result of formatting unstructured data to facilitate storage, use, and generation of information
2. We apply structure (format) based on what we want to perform
    - Some data might not be ready (unstructured) for some type of processing
    - But they may be ready for another type

Example

1. Data: 35116 - Might represent
    - Zip code
    - Sales value
    - Product Code
2. If this value represents a sales value or product code
3. and is stored as text, it cannot be mathematically processed

Another Example

1. Documents stored as images vs. Spreadsheets

Semi-Structured Data

1. has already been processed to some extent
2. Typical web page - has been processed to some extent to show data

Focus of DB

1. Highly Structured Data
2. Corporations use the three classes of data
3. XML DB (Extensible Markup Language)
4. XML DB: supports the storage and management of semistructured XML data

Types of Databases

Definitions

Single User Database

1. Database that supports only one user at a time

Desktop Database

1. Single-user DB that runs on a personal computer

Multiuser DB DB

1. DB that supports multiple concurrent users

Workgroup DB

1. Multiuser DB that usually supports fewer than 50 users
2. or is used for a specific department in an organization

Enterprise DB

1. The overall company data representation
2. Provides support for present and expected future needs

Centralized DB

1. DB located at a single site

Distributed DB

1. Logically related DB
2. that is stored in two or more physically independent sites

Cloud DB

1. DB that is created and maintained using cloud services
2. Cloud services, such as Microsoft Azure, Amazon AWS

General Purpose DB

1. DB that contains a wide variety of data used in multiple disciplines

Discipline-specific DB

1. DB that contains data focused on specific subject areas

Operational DB

1. DB designed primarily to support a company's day-to-day operations
2. aka *transactional database*, *OLTP DB*, *production DB*

Analytical DB

1. DB focused primarily on storing historical data and business metrics
2. used for tactical strategic decision making

Data Warehouse

1. Specialized DB that stores historical and aggregated data in a format optimized for decision support

OLAP

1. Online Analytical Processing
2. A set of tools that provide advanced data analysis for retrieving, processing, and modeling data from the data warehouse

Business Intelligence

1. A set of tools and processes used to capture, collect, integrate, store, and analyze data to support business decision making

Unstructured Data

1. Data that exists in its original, raw state; that is in the format in which it was collected

Structured Data

1. Data that has been formatted to facilitate storage, use, and information generation

Semistructured Data

1. Data that has already been processed to some extent

XML

1. Extensible Markup Language
2. A metalanguage used to represent and manipulate data elements
3. Unlike other markup languages, XML permits the manipulation of a documents' data elements

Types of Databases

Summary

Summary of Types of Databases

TYPES OF DATABASES								
PRODUCT	NUMBER OF USERS			DATA LOCATION		DATA USAGE		XML
	SINGLE USER	MULTIUSER						
		WORKGROUP	ENTERPRISE	CENTRALIZED	DISTRIBUTED	OPERATIONAL	ANALYTICAL	
MS Access	X	X		X		X		
MS SQL Server	X*	X	X	X	X	X	X	X
IBM DB2	X*	X	X	X	X	X	X	X
MySQL	X	X	X	X	X	X	X	X
Oracle RDBMS	X*	X	X	X	X	X	X	X

* Vendor offers single-user/personal or Express DBMS versions

Types of Databases

Social Media and NoSQL

Brief Overview

1. More after we study Data Models
2. https://www.youtube.com/playlist?list=PL-ORy-tsbItySInBIbu7vLiJT6Rq1LGj2
3. https://www.slideshare.net/helghareeb/nosql-databases-not-just-a-buzzword
4. Outdated - Updated version soon

DB Design

DB Design

DB Design Importance

Question

1. Isn't Spreadsheets and Files - Enough?

DB Design

1. the activities that focus on the design of the DB structure that will be used to store and manage end-user data
2. DB that meets all user requirements does not just happen; its structure must be designed carefully
3. Good DBMS performs poorly with a badly designed DB

DB Design

Poor Design

DB Poor Design

EMPLOYEE SKILLS CERTIFICATION IN A POOR DESIGN

Why are there blanks in rows 9 and 10?

How to produce an alphabetical listing of employees?

How to count how many employees are certified in Basic Database Manipulation?

Is Basic Database Manipulation the same as Basic DB Manipulation?

What if an employee acquires a fourth certification?

Do we add another column?

ID	ENum	Name	Title	HireDate	Skill1	Skill1Date	Skill2	Skill2Date	Skill3	Skill3Date
1	02345	Brian Oates	DBA	2/14/1997	Basic Database Management	2/14/2004	Advanced Database Management	2/14/2007	Basic Web Design	8/9/2005
2	08273	Marco Bienz	Analyst	7/28/2009	Basic Web Design	3/8/2011	Advance Process Modeling	8/19/2014		
3	06234	Jasmine Patel	Programmer	8/10/2007	Basic Web Design	8/10/2009	Advanced C# programming	8/10/2009	Basic DB manipulation	1/29/2014
4	03373	Franklin Johnson, Jr.	Purchasing Agent	3/15/2004	Advanced Spreadsheets	6/20/2013				
5	13567	Almond, Robert	Analyst	9/30/2014	Basic Process Modeling	9/30/2016	Basic Database Design	6/29/2017		
6	10282	Richardson, Amanda	Clerk	4/11/2013						
7	09382	Susan Mathis	Database Programmer	8/2/2012	Basic DB Design	8/2/2014	Basic Database Manipulation	8/2/2014	Advanced DB Manipulation	5/1/2015
8	14311	Duong, Lee	Programmer	9/1/2016	Basic Web Design	9/1/2018				
9					Master Database Programming					
10					Basic Spreadsheets					
11	09002	Wade Gaither	Clerk	5/20/2012	Advanced Spreadsheets	5/16/2015	Basic Web Design	5/16/2015		
12	13383	Raymond F. Matthews	Programmer	3/12/2014	Basic C# Programming	3/12/2016				
13	08283	Chavez, Juan	Clerk	7/4/2012						
14	04893	Patricia Richards	DBA	6/11/2006	Advanced Database Management	6/11/2008	Advanced Database Manipulation	9/20/2014		
15	13837	Lee, Megan	Programmer	9/29/2015						

Given the Previous DB Design

1. Produce an alphabetical listing of employees based on their last names
2. Determine how many empoloyees are certified in Basic DB Manipulation
3. Redundant storage of items as a solution might lead to
    - Spelling the names differently
    - Basic Database vs. Basic DB
    - Misspelling
4. DB Structure will change by adding more columns when employees have fourth skill

DB Design

Decomposition Effect

Good DB Design

Database name: Ch01_Text

Table name: EMPLOYEE

Employee_ID	Employee_FName	Employee_LName	Employee_HireDate	Employee_Title
02345	Brian	Oates	2/14/1999	DBA
03373	Franklin	Johnson	3/15/2006	Purchasing Agent
04893	Patricia	Richards	6/11/2008	DBA
06234	Jasmine	Patel	8/10/2009	Programmer
08273	Marco	Blanz	7/28/2010	Analyst
09002	Wade	Gaither	5/20/2014	Clerk
09253	Juan	Chavez	7/4/2014	Clerk
09382	Susan	Mathis	8/2/2014	Database Programmer
10382	Amanda	Richardson	4/11/2015	Clerk
13383	Raymond	Matthews	3/12/2016	Programmer
13567	Robert	Almond	9/30/2016	Analyst
13692	Megan	Lee	9/29/2017	Programmer
14311	Lee	Duong	9/1/2018	Programmer

Table name: CERTIFIED

Employee_ID	Skill_ID	Certified_Date
02345	100	2/14/2004
02345	110	8/9/2005
02345	180	2/14/2007
03373	120	6/20/2013
04893	180	6/11/2008
04893	220	9/20/2014
06234	110	8/10/2009
06234	200	8/10/2009
06234	210	1/29/2014
08273	110	3/8/2011
08273	190	8/19/2014
09002	110	5/16/2015
09002	120	5/16/2015
09382	140	8/2/2014
09382	210	8/2/2014
09382	220	5/1/2015
13383	170	3/12/2016
13567	130	9/30/2016
13567	140	5/29/2017
14311	110	9/1/2018

Table name: SKILL

Skill_ID	Skill_Name	Skill_Description
100	Basic Database Management	Create and manage database user accounts.
110	Basic Web Design	Create and maintain HTML and CSS documents.
120	Advanced Spreadsheets	Use of advanced functions, user-defined functions, and macroing.
130	Basic Process Modeling	Create core business process models using standard libraries.
140	Basic Database Design	Create simple data models.
150	Master Database Programming	Create integrated trigger and procedure packages for a distributed environment.
160	Basic Spreadsheets	Create single tab worksheets with basic formulas
170	Basic C# Programming	Create single-tier data aware modules.
180	Advanced Database Management	Manage Database Server Clusters.
190	Advance Process Modeling	Evaluate and Redesign cross-functional internal and external business processes.
200	Advanced C# Programming	Create multi-tier applications using multi-threading
210	Basic Database Manipulation	Create simple data retrieval and manipulation statements in SQL.
220	Advanced Database Manipulation	Use of advanced data manipulation methods for multi-table inserts, set operations, and correlated subqueries.

Decomposition Characteristics

1. Three distinct but related tables
2. Contain same data
3. Can be easily manipulated
4. Can easily view data in different ways and answer simple questions

SQL Challenge

1. Produce an alphabetical listing of employees by last name

Answer

```
SELECT * FROM EMPLOYEE ORDER BY EMPLOYEE_LNAME;
```

DB Design

Definitions

DB Design

1. The process that yields the description of the DB structure and determines the DB components
2. The second phase of DB Life cycle

SQL Challenge

1. Determine how many employees are certified in Basic Data Manipulation

Answer

```
SELECT COUNT(*)FROM SKILL JOIN CERTIFIED ON SKILL.
    SKILL_ID = CERTIFIED.SKILL_ID
WHERE SKILL_NAME = 'Basic Database Manipulation';
```

Question

1. What changes to tables structure are needed to add fourth, fifth, and sixth skills?

Answer

None

Proper DB Design

1. Designer identifies precisely the DB's expected use
2. Emphasize the use of historical and aggregated data
3. DB used in centralized, single-user environment requires a different approach from the one used in a distributed, and multi-user DB

Evolution of File System Data Processing

Evolution of File System Data Processing

Manual File Systems

Manual File Systems

1. Organizations must develop systems for handling core business tasks
2. Such systems were often manual, paper-and-pencil systems
3. File folders and filing cabinets
4. Collection of data was relatively small and users had few reporting requirements
5. Organizations grew and reporting requirements became complex

Evolution of File System Data Processing

Computerized File Systems

Computerized File Systems

1. Generating reports from manual file systems was slow and cumbersome [18]
2. Weeks of intensive effort each quarter
3. **Data Processing (DP) Specialist**: the person responsible for developing and managing a computerized file processing system

[18] large or heavy and therefore difficult to carry or use; unwieldy

Evolution of File System Data Processing

Definitions

Data Processing (DP) Specialist

1. The person responsible for developing and managing a computerized file processing system

Contents of the CUSTOMER file

C_NAME	C_PHONE	C_ADDRESS	C_ZIP	A_NAME	A_PHONE	TP	AMT	REN
Alfred A. Ramas	615-844-2573	218 Fork Rd., Babs, TN	36123	Leah F. Hahn	615-882-1244	T1	100.00	05-Apr-2018
Leona K. Dunne	713-894-1238	Box 12A, Fox, KY	25246	Alex B. Alby	713-228-1249	T1	250.00	16-Jun-2018
Kathy W. Smith	615-894-2285	125 Oak Ln, Babs, TN	36123	Leah F. Hahn	615-882-2144	S2	150.00	29-Jan-2019
Paul F. Olowski	615-894-2180	217 Lee Ln., Babs, TN	36123	Leah F. Hahn	615-882-1244	S1	300.00	14-Oct-2018
Myron Orlando	615-222-1672	Box 111, New, TN	36155	Alex B. Alby	713-228-1249	T1	100.00	28-Dec-2018
Amy B. O'Brian	713-442-3381	387 Troll Dr., Fox, KY	25246	John T. Okon	615-123-5589	T2	850.00	22-Sep-2018
James G. Brown	615-297-1228	21 Tye Rd., Nash, TN	37118	Leah F. Hahn	615-882-1244	S1	120.00	25-Mar-2019
George Williams	615-290-2556	155 Maple, Nash, TN	37119	John T. Okon	615-123-5589	S1	250.00	17-Jul-2018
Anne G. Farriss	713-382-7185	2119 Elm, Crew, KY	25432	Alex B. Alby	713-228-1249	T2	100.00	03-Dec-2018
Olette K. Smith	615-297-3809	2782 Main, Nash, TN	37118	John T. Okon	615-123-5589	S2	500.00	14-Mar-2019

C_NAME = Customer name
C_PHONE = Customer phone
C_ADDRESS = Customer address
C_ZIP = Customer zip code

A_NAME = Agent name
A_PHONE = Agent phone
TP = Insurance type
AMT = Insurance policy amount, in thousands of $
REN = Insurance renewal date

Basic File Terminology in DB

BASIC FILE TERMINOLOGY	
TERM	**DEFINITION**
Data	Raw facts, such as a telephone number, a birth date, a customer name, and a year-to-date (YTD) sales value. Data has little meaning unless it has been organized in some logical manner.
Field	A character or group of characters (alphabetic or numeric) that has a specific meaning. A field is used to define and store data.
Record	A logically connected set of one or more fields that describes a person, place, or thing. For example, the fields that constitute a record for a customer might consist of the customer's name, address, phone number, date of birth, credit limit, and unpaid balance.
File	A collection of related records. For example, a file might contain data about the students currently enrolled at Gigantic University.

Challenge

1. Apply the acronyms of File Terminology to CUSTOMER file

Evolution of File System Data Processing

File System Redux: Modern End-User Productivity Tools

Modern End-User Productivity Tools

1. Spread of Personal PCs and Spreadsheets
2. The need to access data personally and immediately
3. Gave the illusion of: We can create and manage our own data
4. Using simple tools - Excel sheets and Files

Evolution of File System Data Processing

Definitions

Field

1. A character or group of characters (alphabetic or numeric) that has a specific meaning.
2. A field is used ot define and store data

Record

1. A logically connected set of one or more fields that describes a person, place, or thing

File

1. A collection of related records
2. Example, a file might contain data about the students currently enrolled at Mansoura University

Problems with File System Data Processing

Critiques

1. File systems are better than manual files and served well for over two decades
2. Understanding the shortcomings of the file system enables you to understand the development of modern databases.
3. Failure to understand such problems is likely to lead to their duplication in a database environment, even though database technology makes it easy to avoid them.
4. DBMSs themselves are no gain over file systems

File Systems DB Processing Challenges

1. *Lengthy development times* - Simplest retrieval task requires extensive programming - procedural programming
2. *Difficulty of getting quick answers* - ad hoc queries impossible
3. *Complex System Administration*
4. *Lack of security and limited data sharing*
5. *Extensive Programming*
6. Those challenges lead to: Problems of structural and data dependence

Problems with File System Data Processing

File Systems DB Processing Problems

File Systems DB Processing Problems

1. Structural and Data Dependence
2. Data Redundancy
3. Data Anomalies

Problems with File System Data Processing

Structural and Data Dependence

Structural Dependence

1. File system exhibits structural dependence, which means that access to a file is dependent on its structure
2. Adding / removing new items to a file would require change to the whole file and the program
3. Even changing the characteristics of data - from integer to decimal

Structural Independence

1. We can change the file structure without affecting the application's ability to access the data

Practical Significance (Structural Dependence vs. Independence)

1. Logical vs. Physical Data Format
2. Logical Data Format: how the humans view the data
3. Physical Data Format: how the computer must work with the data

Problems with File System Data Processing

Data Redundancy

Islands of Information

1. File system's structure makes it difficult to combine data from multiple sources
2. Vulnerable to security breaches
3. Organizational structure promotes the storage of the same basic data in different locations
4. Islands of information = scattered data locations
5. Islands of information contain different versions of the same data

Data Redundancy

1. exists when the same data is stored unnecessarily in different places
2. Is Data Redundancy itself a Problem?

Answer

1. No, Controlled Data Redundancy is not a problem
2. Uncontrolled Data Redundancy sets the stage for the following
   - Poor data security
   - Data inconsistency: different and conflicting versions of the same data appear in different places
   - Data-entry errors
   - Data integrity problems: enter non-existing entities (employees, managers, etc.)

Data Integrity

1. Data that displays data inconsistency is also referred to as data that lacks data integrity.
2. Data integrity is defined as the condition in which all of the data in the database is consistent with the real-world events and conditions.
3. Data integrity means the following:
   3.1 Data is accurate: there are no data inconsistencies.
   3.2 Data is verifiable: the data will always yield consistent results

Problems with File System Data Processing

Data Anomalies

Data Anomaly

1. Change should happen in one place, however Uncontrolled Data Redundancy makes that impossible
2. Data anomaly develops when not all of the required changes in the redundant data are made successfully
    - Update anomalies
    - Insertion anomalies
    - Deletion anomalies

Problems with File System Data Processing

Definitions

Structural Dependence

1. A data characteristic in which a change in the DB schema affects data access, thus requiring changes in all access programs

Structural Independence

1. A data characteristic in which changes in the DB schema do not affect data access

Data Type

1. Defines the kind of values that can be used or stored
2. Also, used in programming languages and DB systems to determine the operations that can be applied to such data

Data Dependence

1. A data condition in which data representation and manipulation are dependent on the physical data storage characteristic

Data Independence

1. A condition in which data access is unaffected by changes in the physical data storage characteristics

Logical Data Format

1. The way a person views data within the context of a problem domain

Physical Data Format

1. The way a computer "sees" (stores) data

Islands of Information

1. In the old file system environment, pools of independent, often duplicated, and inconsistent data created and managed by different department

Data Redundancy

1. Exists when the same data is stored unnecessarily at different places

Data Integrity

1. In a relational DB, a condition in which the data in the DB compiles with all entity and referential integrity constraints

Data Anomaly

1. A data abnormality in which inconsistent changes have been made to a database
2. For example, an employee moves, but the address change is not corrected in all files in the database

Database Systems

File Systems vs. Database Systems

1. File system consists of many separated and unrelated files
2. Database system consists of logically related data stored in a single logical data repository
3. The "logical" label reflects the fact that the data repository appears to be a single unit to the end user, even though data might be physically distributed among multiple storage facilities and locations.

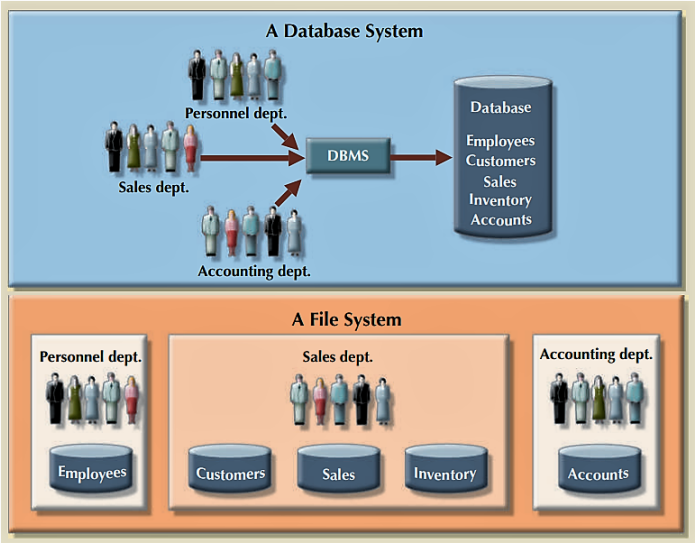
File Systems vs. Database Systems

1. Database Systems eliminates most of the file systems'
    - data inconsistency
    - data anomaly
    - data dependence
    - structural dependence.

Better Yet

1. DBMS stores not only the data structures but also
    - relationships between those structures
    - access paths to those structures
2. All in a central location

File Systems vs. Database Systems



Database Systems

DB Systems Environment

DB System

1. Organization of components that define and regulate the collection, storage, management, and use of data within a database environment
2. From a general management point of view, the database system is composed of the five major parts
   2.1 Hardware
   2.2 Software
   2.3 People
   2.4 Procedures
   2.5 Data

Hardware

1. Refers to all of the system's physical devices, including
    - Computers (PCs, tablets, workstations, servers, and supercomputers)
    - Storage devices
    - Printers
    - Network devices (hubs, switches, routers, fiber optics)
    - Other devices (automated teller machines, ID readers, and so on)

Software - 01

1. Three types of software are needed: operating system software, DBMS software, and application programs and utilities.
   1.1 **Operating system software** manages all hardware components and makes it possible for all other software to run on the computers
   1.2 **DBMS software** manages the database within the database system

Software - 02

3. **Application programs and utility software** are used to access and manipulate data in the DBMS and to manage the computer environment in which data access and manipulation take place
   - Most commonly used to access data within the database to generate reports, tabulations, and other information to facilitate decision making
   - **Utilities** are the software tools used to help manage the database system's computer components. For example, Graphical User Interfaces (GUIs) to help create database structures, control database access, and monitor database operations

People

1. All users of the database system
2. Five types of users can be identified:
   2.1 System Administrators
   2.2 Database Administrators
   2.3 Database Designers
   2.4 System Analysts and Programmers
   2.5 End Users
3. Each user type, performs both unique and complementary functions.

System Administrators

1. Oversee the database system's general operations

Database Administrators

1. DBAs
2. Manage the DBMS and ensure that the database is functioning properly
3. Database Administration and Security

Database Designers

1. Design the database structure
2. Database Architects
3. If the database design is poor, even the best application programmers and the most dedicated DBAs cannot produce a useful database environment
4. Organizations strive to optimize their data resources, the database designer's job description has expanded to cover new dimensions and growing responsibilities.

System Analysts and Programmers

1. Design and implement the application programs
2. They design and create the data-entry screens, reports, and procedures through which end users access and manipulate the database's data

End Users

1. The people who use the application programs to run the organization's daily operations
2. For example, sales clerks, supervisors, managers, and directors are all classified as end users
3. High-level end users employ the information obtained from the database to make tactical and strategic business decisions

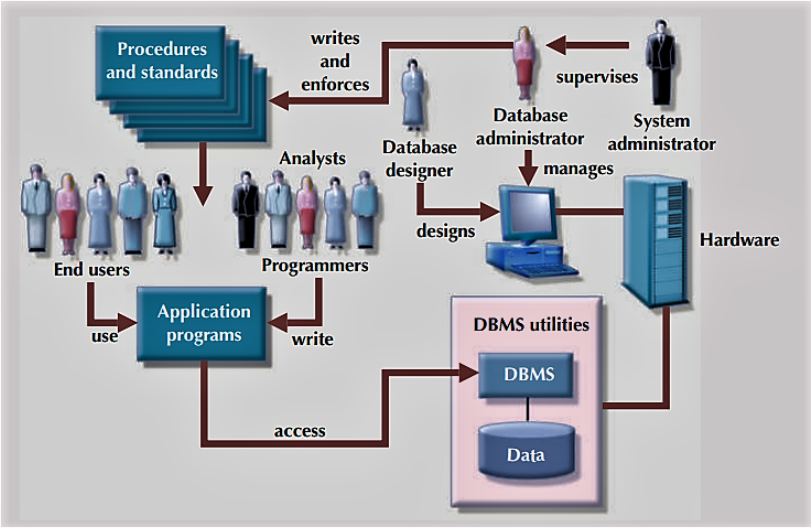
Procedures

1. Instructions and rules that govern the design and use of the database system
2. Critical, although occasionally forgotten, component of the system
3. Enforce the standards by which business is conducted within the organization and with customers
4. Ensure that companies have an organized way to monitor and audit the data that enter the database and the information generated from those data

Data

1. the collection of facts stored in the database
2. determining which data to enters into the database and how to organize that data is a vital part of the database designer's job

DB Systems Environment



Database Systems

Definitions

Database System

1. An organization of components that defines and regulates the collection, storage, management, and use of data in a database environment

DBMS Functions

DBMS Functions

1. Guarantee integrity and consistency of data in database
2. Most of those functions are transparent to end users
3. They include
   3.1 Data dictionary management
   3.2 Data storage management
   3.3 Data transformation and presentation
   3.4 Security management
   3.5 Multiuser access control
   3.6 Backup and recovery management
   3.7 Data integrity management
   3.8 Database access languages and application programming interfaces
   3.9 Database communication interfaces

DBMS Functions

1: Data Dictionary Management

Data Dictionary Management

1. DBMS stores definitions of the data elements and their relationships (metadata) in a data dictionary
2. All programs that access the data in the database work through the DBMS
3. DBMS uses the data dictionary to look up the required data component structures and relationships
4. Changes made in a database structure are automatically recorded in the data dictionary
5. In other words, the DBMS provides data abstraction, and it removes structural and data dependence from the system

DBMS Functions

2: Data Storage Management

Data Storage Management - 01

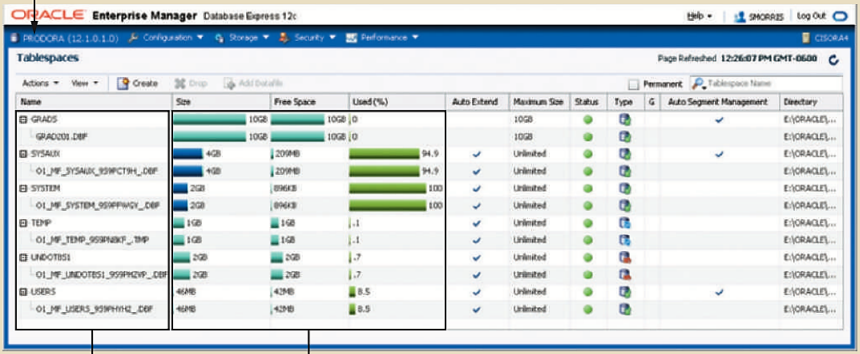
1. DBMS creates and manages the complex structures required for data storage
2. Modern DBMS provides storage not only for the data but also for related data-entry forms or screen definitions, report definitions, data validation rules, procedural code, structures to handle video
3. Although the user sees the database as a single data storage unit, the DBMS actually stores the database in multiple physical data files and picture formats, and so on
4. **Database performance tuning**: Performance tuning relates to the activities that make the database perform more efficiently in terms of storage and access speed

Data Storage Management - 02

5. Such data files may even be stored on different storage media
6. Therefore, the DBMS does not have to wait for one disk request to finish before the next one starts.
7. DBMS can fulfill database requests concurrently

Data Storage Management Example

Database Name: PRODORA



The PRODORA database is actually stored in six physical datafiles organized into six logical tablespaces located on the E: drive of the database server computer

The Oracle Enterprise Manager Express interface also shows the amount of space used by each of the datafiles.

Activate Wi
Go to Settings t

DBMS Functions

3: Data Transformation and Presentation

Data Transformation and Presentation

1. DBMS transforms entered data to conform to required data structures
2. DBMS formats the physically retrieved data to make it conform to the user's logical expectations
   - An end user in Egypt would expect to enter the date July 11, 2017, as "11/07/2017."
   - In contrast, the same date would be entered in the United States as "07/11/2017."
   - Regardless of the data presentation format, the DBMS must manage the date in the proper format for each country

DBMS Functions

4: Security Management

Security Management

1. The DBMS creates a security system that enforces user security and data privacy
2. Security rules determine
   - which users can access the database
   - which data items each user can access
   - which data operations (read, add, delete, or modify) the user can perform
3. Important in multiuser database systems
4. DB users may be authenticated to the DBMS through a username and password or through biometric authentication such as a fingerprint scan
5. DBMS uses this information to assign access privileges to various database components such as queries and reports

DBMS Functions

5: Multiuser Access Control

Multiuser Access Control

1. To provide data integrity and data consistency
2. DBMS uses sophisticated algorithms to ensure that multiple users can access the database concurrently without compromising its integrity

DBMS Functions

6: Backup and Recovery Management

Backup and Recovery Management

1. DBMS provides backup and data recovery to ensure data safety and integrity
2. DBMS systems provide special utilities that allow the DBA to perform routine and special backup and restore procedures
3. Recovery management deals with the recovery of the database after a failure, such as a bad sector in the disk or a power failure
4. Such capability is critical to preserving the database's integrity

DBMS Functions

7: Data Integrity Management

Data Integrity Management

1. DBMS promotes and enforces integrity rules
2. Minimizing data redundancy and Maximizing data consistency
3. Data relationships stored in the data dictionary are used to enforce data integrity
4. Important in transaction-oriented database systems

DBMS Functions

8: DB Access Languages and APIs

Database Access Languages and Application Programming Interfaces (APIs)

1. DBMS provides data access through a query language
2. Query language is a nonprocedural language
3. Query Language lets the user specify what must be done without having to specify how
4. Query Language vs. Programming Language
5. Structured Query Language (SQL) is the *de facto* query language and data access standard supported by the majority of DBMS vendors
6. DBMS provides APIs to programming languages (Python, C♯)
7. DBMS provides administrative utilities used by the DBA and the database designer to create, implement, monitor, and maintain the database

DBMS Functions

9: Database Communication Interfaces

Database Communication Interfaces

1. DBMS accepts end-user requests via multiple, different network environments
2. DBMS might provide access to the database via the Internet through the use of web browsers
3. Communications can be accomplished in several ways
   3.1 End users can generate answers to queries by filling in screen forms through their preferred web browser
   3.2 DBMS can automatically publish predefined reports on a website
   3.3 DBMS can connect to third-party systems to distribute information via email or other productivity applications

DBMS Functions

Question

Why a Spreadsheet Is Not a Database?

1. Spreadsheet allows for the manipulation of data in a tabular format
2. It does not support even the most basic database functionality such as
   2.1 support for self-documentation through metadata
   2.2 enforcement of data types or domains to ensure consistency of data within a column
   2.3 defined relationships among tables
   2.4 constraints to ensure consistency of data across related tables
3. Most users lack the necessary training to recognize the limitations of spreadsheets for these types of tasks

DBMS Functions

Definitions

Data Dictionary

1. A DBMS component that stores metadata - data about data.
2. The data dictionary contains data definitions as well as data characteristics and relationships
3. May also include data that is external to the DBMS

Performance Tuning

1. Activities that make a database perform more efficiently in terms of storage and access speed

Query Language

1. A nonprocedural language that is used by a DBMS to manipulate its data
2. An example of a query language is SQL

Structured Query Language (SQL)

1. A powerful and flexible relational database language composed of
   1.1 commands that enable users to
      - create database and table structures
      - perform various types of data manipulation and data administration
      - query the database to extract useful information

DBMS Disadvantages

DBMS Disadvantages

1. *Increased costs* - HW, SW, Training, Licenses, Management, etc.
2. *Management Complexity*
3. *Maintaining Currency* - frequent updates, latest patches, personal training, etc.
4. *Vendor Dependence*
5. *Frequent upgrade/replacement cycles*

DB Professional Career

DB Professional Career

Career Opportunities

Career Opportunities

JOB TITLE	DESCRIPTION	SAMPLE SKILLS REQUIRED
Database Developer	Create and maintain database-based applications	Programming, database fundamentals, SQL
Database Designer	Design and maintain databases	Systems design, database design, SQL
Database Administrator	Manage and maintain DBMS and databases	Database fundamentals, SQL, vendor courses
Database Analyst	Develop databases for decision support reporting	SQL, query optimization, data warehouses
Database Architect	Design and implementation of database environments (conceptual, logical, and physical)	DBMS fundamentals, data modeling, SQL, hardware knowledge, etc.
Database Consultant	Help companies leverage database technologies to improve business processes and achieve specific goals	Database fundamentals, data modeling, database design, SQL, DBMS, hardware, vendor-specific technologies, etc.
Database Security Officer	Implement security policies for data administration	DBMS fundamentals, database administration, SQL, data security technologies, etc.
Cloud Computing Data Architect	Design and implement the infrastructure for next-generation cloud database systems	Internet technologies, cloud storage technologies, data security, performance tuning, large databases, etc.
Data Scientist	Analyze large amounts of varied data to generate insights, relationships, and predictable behaviors	Data analysis, statistics, advanced mathematics, SQL, programming, data mining, machine learning, data visualization

Go to Settings to activate Wi

DB Professional Career

Advanced DB Technologies

Very Large Databases (VLDB)

1. DBs that support large amounts of data, usually in the petabyte range [19]
2. VLDB vendors include Oracle Exadata, IBM's Netezza, HP's Vertica, and Teradata
3. VLDBs are now being superceded by Big Data databases
4. NoSQL Generally
5. Products such as Cassandra (Facebook) and BigTable (Google) are using "columnar-database" technologies

[19] A petabyte is more than 1,000 terabytes

In-Memory Databases

1. Address the need for faster database processing
2. Store most of their data in primary memory (RAM) rather than in slower secondary storage (hard disks)
3. In-memory databases include IBM's solidDB and Oracle's TimesTen

Cloud Databases

1. Cloud database services allow organizations to quickly add database systems to their environment
2. While simultaneously lowering the total cost of ownership of a new DBMS (servers, licenses, etc.)
3. Offers all the advantages of a local DBMS
4. Instead of residing within organization's network infrastructure, it resides on the Internet

Summary

Summary (1/8)

1. Data consists of raw facts
2. Information is the result of processing data to reveal its meaning
3. Accurate, relevant, and timely information is the key to good decision making
4. Good decision making is the key to organizational survival in a global environment

Summary (2/8)

5. Data is usually stored in a database
6. To implement a database and to manage its contents, you need a DataBase Management System (DBMS)
7. DBMS serves as the intermediary between the user and the database
8. DB contains the data you have collected and "data about data," known as metadata

Summary (3/8)

9. Database design defines the database structure
10. A well-designed database facilitates data management and generates accurate and valuable information
11. A poorly designed database can lead to poor decision making, and poor decision making can lead to the failure of an organization

Summary (4/8)

12. Databases can be classified according to the number of users supported, where the data is located, the type of data stored, the intended data usage, and the degree to which the data is structured

Summary (5/8)

13. Databases evolved from manual and then computerized file systems.
14. In a file system, data is stored in independent files, each requiring its own data management programs.
15. Although this method of data management is largely outmoded, understanding its characteristics makes database design easier to comprehend.

Summary (6/8)

16. Some limitations of file system data management are that it requires extensive programming, system administration can be complex and difficult, making changes to existing structures is difficult, and security features are likely to be inadequate
17. Also, independent files tend to contain redundant data, leading to problems of structural and data dependence.

Summary (7/8)

18. DBMSs were developed to address the file system's inherent weaknesses
19. Rather than depositing data in independent files, a DBMS presents the database to the end user as a single data repository. This arrangement promotes data sharing, thus eliminating the potential problem of islands of information
20. In addition, the DBMS enforces data integrity, eliminates redundancy, and promotes data security

Summary (8/8)

21. Knowledge of database technologies leads to many career opportunities in the ever-expanding IT industry
22. There is a variety of specialization within the database arena for a wide range of skills and expertise

https://www.haitham.ws