

Key Points

- the components of a database system
- the advantages of databases over files
- types of database systems
- relational databases

Components of a Database System

- database
 - the data – a structured collection of information
- *structured data* is highly organized and easily searchable
 - fits neatly into data tables with rows and columns
 - conforms to a defined data model or schema which specifies the format and types of data for each column
- *unstructured data* lacks a defined data model or schema
 - does not fit neatly into data tables
 - e.g. text, images, videos, documents

Components of a Database System

- database
 - the data – a structured collection of information
- database management system (DBMS)
 - the system providing all the services governing access to the data – efficient querying, integrity constraints, access control, concurrency, backups and recovery, ...
- application programs
 - programs that use the data via interactions with the DBMS

Advantages of Databases Over Files

- centralized storage of data avoids redundancy and inconsistency across different files
- data isn't isolated in an obscure file
- self-describing
 - contains both data and metadata defining and describing the data
- program-data independence
 - refers to separation between application program and the data
 - application programs work with a conceptual model of the data provided by the DBMS instead of directly with the underlying files storing the data
 - allows the storage details to change without affecting the application program

Advantages of Databases Over Files

- added value
 - integrity and consistency constraints are part of the database and enforced by the DBMS
 - security, concurrency, transactions, backup and recovery are provided by the DBMS
 - *transactions* refer to bundling a multi-step operation into a single unit that either fully completes or doesn't happen at all
 - ...though these services do incur overhead (disk space, time)

Types of Database Systems

- single-user vs multiuser
 - *single-user* systems are typically limited to personal computers
 - e.g. a back end for another program such as Microsoft Outlook
 - still useful because DBMS provides other features besides concurrency control (efficient querying, backup and crash recovery, etc)

Types of Database Systems

- centralized vs distributed
 - *centralized* exists on a single server, *distributed* involves multiple servers (for mirroring or different parts of the database)
 - distributed can be further classified as *homogeneous* or *heterogeneous* depending on whether the same DBMS software is used at all of the sites
 - heterogeneous can occur when different organizations maintain databases that are integrated into a single package
 - e.g. a library catalog that includes the holdings of many different libraries
 - e.g. data clearinghouses

Types of Database Systems

- data model – relational vs other things
 - relational model is based on tables (with rows and columns)
 - most common
 - general purpose
 - object-oriented database stores objects (instance variables and methods)

Relational Databases

- a relational database consists of tables
- each row in a table groups together data values about a particular thing
 - duplicate rows are not allowed (must differ by at least one value)
 - order of the rows doesn't matter
 - order of the columns doesn't matter
 - data values are *atomic*
- foreign keys link tables together
 - e.g. DEPARTMENT.Mgr_ssn → EMPLOYEE.Ssn

Empno	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fordren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1965-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alice	J	Zelaysa	999887777	1968-01-19	3321 Castra, Spring, TX	F	25000	987654321	4
Jennifer	S	Walshca	987654321	1941-08-20	291 Berny, Belare, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1982-09-15	975 Five Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5831 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	50000	NULL	1

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

Essn	Pno	Hours	Phname	Pnumber	Plocation	Dnum
123456789	1	32.5	ProductX	1	Belare	5
123456789	2	3.5	ProductY	2	Sugarland	5
666884444	3	40.0	ProductZ	3	Houston	5
453453453	1	20.0	Computerization	10	Stafford	4
453453453	2	20.0	Reorganization	20	Houston	1
333445555	2	10.0	Newbenefits	30	Stafford	4
333445555	3	10.0				
333445555	10	10.0				

Essn	Dependent_name	Sex	Bdate	Relationship			
999887777	333445555	Alice	F	1988-04-05	Daughter		
999887777	10	10.0	333445555	Theodore	M	1983-10-25	Son
987987987	10	35.0	333445555	Joy	F	1958-05-03	Spouse
987987987	30	5.0	987654321	Abner	M	1942-02-28	Spouse
987654321	30	20.0	123456789	Michael	M	1988-01-04	Son
987654321	20	15.0	123456789	Alice	F	1988-12-30	Daughter
888665555	20	NULL	123456789	Elisabeth	F	1987-05-05	Spouse

Relational Databases

- terminology

theoretical	SQL	DBMS internals
relation	table	file
tuple	row	record
attribute	column	field

 - domain is the data type for a column
 - a view is a subset of rows and columns from one or more tables
 - a database schema defines the tables, columns, and constraints

Empno	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Forderen, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1965-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alice	J	Zelaysa	999887777	1968-01-19	3321 Castra, Spring, TX	F	25000	987654321	4
Jennifer	S	Walshca	987654321	1941-08-20	291 Berny, Belare, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1982-09-15	975 Five Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5831 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	50000	NULL	1

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

Essn	Pno	Hours	Phname	Pnumber	Plocation	Dnum
123456789	1	32.5	ProductX	1	Belare	5
123456789	2	3.5	ProductY	2	Sugarland	5
666884444	3	40.0	ProductZ	3	Houston	5
453453453	1	20.0	Computerization	10	Stafford	4
453453453	2	20.0	Reorganization	20	Houston	1
333445555	2	10.0	Newbenefits	30	Stafford	4
333445555	3	10.0				
333445555	10	10.0				

Essn	Dependent_name	Sex	Bdate	Relationship			
999887777	333445555	Alice	F	1988-04-05	Daughter		
999887777	10	10.0	333445555	Theodore	M	1983-10-25	Son
987987987	10	35.0	333445555	Joy	F	1958-05-03	Spouse
987987987	30	5.0	987654321	Abner	M	1942-02-28	Spouse
987654321	30	20.0	123456789	Michael	M	1988-01-04	Son
987654321	20	15.0	123456789	Alice	F	1988-12-30	Daughter
888665555	20	NULL	123456789	Elisabeth	F	1987-05-05	Spouse

Relational Databases

- notation for relational schemas
- EMPLOYEE (Fname, Minit, Lname, Ssn, Bdate, Address, Sex, Salary, Super_ssn, Dno)
- DEPARTMENT (Dname, Dnumber, Mgr_ssn, Mgr_start_date)
- WORKS_ON (Essn, Pno, Hours)
- ...
- underline denotes the primary key – two rows cannot have the same values for all the fields of the PK
 - PK uniquely identifies each row
 - foreign keys might be shown with arrows or listed separately
 - DEPARTMENT.Mgr_ssn → EMPLOYEE.Ssn
 - EMPLOYEE.Super_ssn → EMPLOYEE.Ssn
 - WORKS_ON.Essn → EMPLOYEE.Ssn
 - EMPLOYEE.Dno → DEPARTMENT.Dnumber

Empno	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Forderen, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1965-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alice	J	Zelaysa	999887777	1968-01-19	3321 Castra, Spring, TX	F	25000	987654321	4
Jennifer	S	Walshca	987654321	1941-08-20	291 Berny, Belare, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1982-09-15	975 Five Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5831 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	50000	NULL	1

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

Essn	Pno	Hours	Phname	Pnumber	Plocation	Dnum
123456789	1	32.5	ProductX	1	Belare	5
123456789	2	3.5	ProductY	2	Sugarland	5
666884444	3	40.0	ProductZ	3	Houston	5
453453453	1	20.0	Computerization	10	Stafford	4
453453453	2	20.0	Reorganization	20	Houston	1
333445555	2	10.0	Newbenefits	30	Stafford	4
333445555	3	10.0				
333445555	10	10.0				

Essn	Dependent_name	Sex	Bdate	Relationship			
999887777	333445555	Alice	F	1988-04-05	Daughter		
999887777	10	10.0	333445555	Theodore	M	1983-10-25	Son
987987987	10	35.0	333445555	Joy	F	1958-05-03	Spouse
987987987	30	5.0	987654321	Abner	M	1942-02-28	Spouse
987654321	30	20.0	123456789	Michael	M	1988-01-04	Son
987654321	20	15.0	123456789	Alice	F	1988-12-30	Daughter
888665555	20	NULL	123456789	Elisabeth	F	1987-05-05	Spouse