

Roadmap – Foundations

Last time –

- why databases?

Today –

- how a database management system is built
 - components
 - database architecture
 - data independence
- the responsibility that comes with all of it
 - data custodianship
- relational vocabulary

Next time –

- querying – SQL

Components of a Database System

- database – a *structured* collection of data
 - *data* – the actual stored information
 - *metadata* – data about the data (schema definitions, constraints, data types)

Structured vs Unstructured Data

- *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
 - unstructured storage of data may be paired with a structured or semi-structured database for the associated metadata
 - e.g. relational database stores title, upload date, filename or URL for videos stored elsewhere
- *semi-structured data* has some organizational structure but not a rigid, predefined schema
 - structure exists but can be nested (fields within other fields) and may vary from one record to the next
 - doesn't fit neatly into data tables but has other mechanisms (such as tags, keys, or markers) to separate and label elements
 - e.g. JSON documents, XML files, NoSQL document stores

Components of a Database System

- database – a *structured* collection of data
 - *data* – the actual stored information
 - *metadata* – data about the data (schema definitions, constraints, data types)
- 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, ...
- users
- application programs
 - programs that use the data via interactions with the DBMS

DB Users

- end users
 - interact with the DBMS through an app
 - database is hidden behind the app's UI
- sophisticated users
 - interact with the DBMS through a client
 - write their own queries
- application programmers
 - write the code (and embedded queries) that end users interact with
- database administrators (DBA)
 - manage the system itself (schema design, security, performance, backups)

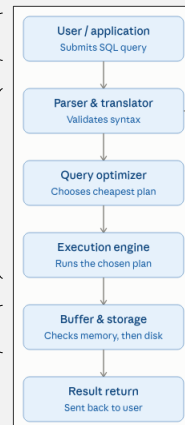
Core DBMS Components and Functions

- query engine – interprets and executes the query
 - *query mechanisms* are how *what* you want (the query) gets translated into *how* to get it (an execution plan)
- transaction manager
 - supports grouping multiple operations so they succeed or fail together as a unit
 - supports concurrent operations while maintaining data integrity
- storage manager – manages how data is stored on the disk
 - *access methods* (or *access paths*) define how the DBMS efficiently locates specific records
 - e.g. scan the whole file, use an index, ...
- buffer management
 - manages transfer of data between memory and disk
 - disks are slow, but memory isn't big enough for all of the data

From Query to Results

"find all orders placed by customer #100 in the last 6 months"
– user writes SQL directly or app translates UI action into SQL

query processor



storage manager retrieves the data for the execution engine
access paths are strategies for how the storage manager locates the right data
the *buffer manager* is part of the storage manager – responsible for managing what data is in memory

Database Architecture

- database structure can be specified at different levels of abstraction
- *three-schema architecture*
 - *physical/internal schema* – how data is actually stored on the disk
 - file organization, indexing structures, storage details
 - *logical/conceptual schema* – what data exists and how it is related
 - e.g. entities, attributes, relationships
 - *view/external schema* – how a particular user or application sees the data
 - most users only ever need knowledge of the view level

Data Independence

- changing one aspect of database structure shouldn't have a ripple effect requiring changes elsewhere
- physical data independence**
 - can change how data is stored without changing the conceptual schema
 - e.g. add an index on last name to speed up queries filtering by student last name – adds a new access path but still have the same tables
 - e.g. migrate from one physical storage engine to another, changing how records are stored in files or how the files are laid out on disk – only affects the storage manager
- logical data independence**
 - can change the conceptual schema without changing the applications built on top
 - e.g. add a "preferred name" field to the Students table – existing reports and applications that don't use this field shouldn't be affected
- data independence is a (good) consequence of abstraction and the three-schema architecture

33

Data Custodianship Preview

- whoever manages data bears ongoing responsibility for it
 - must keep the data accurate and available
 - must handle the data appropriately on behalf of the people it describes
- involves technical, ethical, and social dimensions
 - technical
 - enforcing access control so only authorized users can view or modify data
 - properly destroying data when it is no longer needed or permitted to be kept
 - ethical
 - using data only for the purposes it was originally collected for
 - being honest with people about what's collected and how it's used
 - social
 - recognizing that a data breach or misuse can affect entire communities (not just the individuals named in the data) and can disproportionately harm marginalized or vulnerable groups

CPSC 343: Database Theory and Practice • Fall 2026

34

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)
 - ...

CPSC 343: Database Theory and Practice • Fall 2026

37

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)
- primary key** – column(s) which uniquely identify rows
 - two rows cannot have the same values for all the fields of the PK
- foreign key** – references a row in another table
 - e.g. DEPARTMENT.Mgr_ssn → EMPLOYEE.Ssn

EMPLOYEE									
Empid	Empname	Dept	Job	Salary	Super_empid	Start_date	End_date	Address	Phone
123456789	John B Smith	123456789	1985-01-08	751	Ford	Houston, TX	M	30000	333445555
987654321	T Wong	333445555	1985-12-08	100	Wong	Houston, TX	M	40000	987654321
456789012	J Zhang	987654321	1986-01-19	200	Cole	Spring, TX	F	25000	987654321
333445555	S Wallace	987654321	1981-06-20	291	Berry	Bellevue, TX	F	43000	888888888
888888888	K Nguyen	987654321	1982-08-10	275	Fin	Dallas, TX	M	38000	333445555
987654321	A Taylor	456789012	1982-01-31	180	Fin	Houston, TX	F	25000	333445555
987654321	V Jaffer	987654321	1989-02-28	280	Dallas	Houston, TX	M	25000	987654321
888888888	E Borg	888888888	1987-11-10	450	Strom	Houston, TX	M	55000	NULL

DEPARTMENT			
Deptname	Deptid	Mgr_ssn	Mgr_start_date
Research	9	333445555	1988-09-22
Administration	4	987654321	1985-01-01
Headquarters	1	888888888	1981-06-19

DEPT LOCATIONS			
Deptname	Location	Address	City
1	Houston		
4	Stafford		
5	Bellevue		
5	Stafford		
5	Houston		

WORKS ON			
Empid	Projid	Hours	Start_date
123456789	1	30.5	
123456789	2	75	
987654321	3	40.0	
456789012	1	20.0	
456789012	2	20.0	
333445555	3	10.0	
333445555	3	10.0	
333445555	10	10.0	
333445555	20	10.0	
999887777	30	30.0	
999887777	10	10.0	
987654321	10	30.0	
987654321	30	5.0	
987654321	30	30.0	
888888888	30	10.0	

PROJECT			
Projid	Projname	Location	Deptid
ProductX	1	Bellevue	5
ProductY	2	Stafford	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newdevelops	30	Stafford	4

DEPENDENT				
Empid	Dependent_name	Sex	Birthdate	Relationship
333445555	Alice	F	1986-04-05	Daughter
999887777	Theodore	M	1983-10-25	Son
987654321	Jay	F	1988-09-09	Son
987654321	Alice	M	1993-02-28	Son
888888888	Michael	M	1988-01-04	Son
987654321	Alice	F	1988-12-30	Daughter
888888888	Elizabeth	F	1987-09-05	Son

CPSC 343: Database Theory and Practice • Fall 2026

39

Relational Databases

terminology

theoretical	SQL
relation	table
tuple	row
attribute	column

- a *database schema* defines the tables, columns, and constraints

EMPLOYEE									
Empid	Empname	Empname	Empname	Empname	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1980-01-09	731 Redwood, Houston, TX	M	20000	333445555	5
Franklin	T	Wong	333445555	1985-12-08	838 Vine, Houston, TX	M	40000	888865555	5
Alice	J	Zelaya	999887777	1988-01-19	3321 Cresta, Spring, TX	F	25000	987654321	4
Jennifer	S	Wakana	987654321	1941-06-20	281 Berry, Dallas, TX	F	40000	888865555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Pine Oak, Humble, TX	M	38000	333445555	5
Jayce	A	English	454345433	1972-07-31	5631 Rice, Houston, TX	F	20000	333445555	5
Armed	V	Jachar	987657987	1989-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
Jane	E	Bray	888665555	1927-11-10	400 Stone, Houston, TX	M	50000	NULL	1

DEPARTMENT				DEPT LOCATIONS	
Dname	Dnumber	Mgr_ssn	Mgr_start_date	Dnumber	Location
Research	5	333445555	1988-05-22	1	Houston
Administration	4	987654321	1989-01-01	4	Stafford
Headquarters	1	888665555	1981-08-19	5	Bakers
				5	Bugland
				5	Houston

WORKS_ON				PROJECT			
Empid	Pno	Hours		Pname	Pnumber	Plocation	Dnumber
123456789	1	32.5		ProductX	1	Bakers	5
123456789	2	7.5		ProductY	2	Stafford	5
666884444	3	40.0		ProductZ	3	Houston	5
454345433	1	20.0		Computation	10	Stafford	4
454345433	5	20.0		Recognition	20	Houston	1
333445555	2	10.0		Newbenefit	30	Stafford	4
333445555	3	10.0					
333445555	10	10.0					
333445555	20	10.0					
999887777	30	30.0					
999887777	10	10.0					
987657987	10	30.0					
987657987	30	8.0					
987654321	30	20.0					
987654321	20	15.0					
888665555	20	NULL					

DEPENDENT					
Empid	Dependent_name	Sex	Bdate	Relationship	
333445555	Alice	F	1986-04-05	Daughter	
999887777	Theodore	M	1983-10-25	Son	
987657987	Jay	F	1954-06-03	Spouse	
987657987	Almar	M	1942-02-28	Spouse	
987654321	Michael	M	1988-01-04	Son	
987654321	Alice	F	1988-12-30	Daughter	
888665555	Elizabeth	F	1967-05-05	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

EMPLOYEE									
Empid	Empname	Empname	Empname	Empname	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1980-01-09	731 Redwood, Houston, TX	M	20000	333445555	5
Franklin	T	Wong	333445555	1985-12-08	838 Vine, Houston, TX	M	40000	888865555	5
Alice	J	Zelaya	999887777	1988-01-19	3321 Cresta, Spring, TX	F	25000	987654321	4
Jennifer	S	Wakana	987654321	1941-06-20	281 Berry, Dallas, TX	F	40000	888865555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Pine Oak, Humble, TX	M	38000	333445555	5
Jayce	A	English	454345433	1972-07-31	5631 Rice, Houston, TX	F	20000	333445555	5
Armed	V	Jachar	987657987	1989-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
Jane	E	Bray	888665555	1927-11-10	400 Stone, Houston, TX	M	50000	NULL	1

DEPARTMENT				DEPT LOCATIONS	
Dname	Dnumber	Mgr_ssn	Mgr_start_date	Dnumber	Location
Research	5	333445555	1988-05-22	1	Houston
Administration	4	987654321	1989-01-01	4	Stafford
Headquarters	1	888665555	1981-08-19	5	Bakers
				5	Bugland
				5	Houston

WORKS_ON				PROJECT			
Empid	Pno	Hours		Pname	Pnumber	Plocation	Dnumber
123456789	1	32.5		ProductX	1	Bakers	5
123456789	2	7.5		ProductY	2	Stafford	5
666884444	3	40.0		ProductZ	3	Houston	5
454345433	1	20.0		Computation	10	Stafford	4
454345433	5	20.0		Recognition	20	Houston	1
333445555	2	10.0		Newbenefit	30	Stafford	4
333445555	3	10.0					
333445555	10	10.0					
333445555	20	10.0					
999887777	30	30.0					
999887777	10	10.0					
987657987	10	30.0					
987657987	30	8.0					
987654321	30	20.0					
987654321	20	15.0					
888665555	20	NULL					

DEPENDENT					
Empid	Dependent_name	Sex	Bdate	Relationship	
333445555	Alice	F	1986-04-05	Daughter	
999887777	Theodore	M	1983-10-25	Son	
987657987	Jay	F	1954-06-03	Spouse	
987657987	Almar	M	1942-02-28	Spouse	
987654321	Michael	M	1988-01-04	Son	
987654321	Alice	F	1988-12-30	Daughter	
888665555	Elizabeth	F	1967-05-05	Spouse	