Bases de données

Lecture 1

Nabil Mustafa

mustafa@lipn.univ-paris13.fr

Storing Data

Goal: store, maintain, search data

- USPN: records of all students, teachers, grades, courses
- Darty: products and all their information
- Songs: artist, year of release, type of song, album cover
- Banking: transactions with user accounts

Need to be able to:

update (add, delete, modify)

submit new grades of all students for the module `Databases`

search (called **querying**)

all teachers who taught C++ programming

Storing Data

First try: store pictures, screenshots of everything as image files on a computer

+ flexible, lots of information
- rigid, difficult to search/manipulate

Second try: a list of files (`doc`, `pdf`, `xls`, `txt`, ...) on your computer

+ easier to access data
- must write programs to search
- duplicate or inconsistent information
- multiple formats

Major Concerns: Scalability, Security, Efficiency, Concurrency, . . .

Storing Data

For wide use, **must** satisfy *many* criteria:

fast search

many concurrent users

consistency when modifying data

secure from unauthorized access

be able to recover from errors

always correct search reply

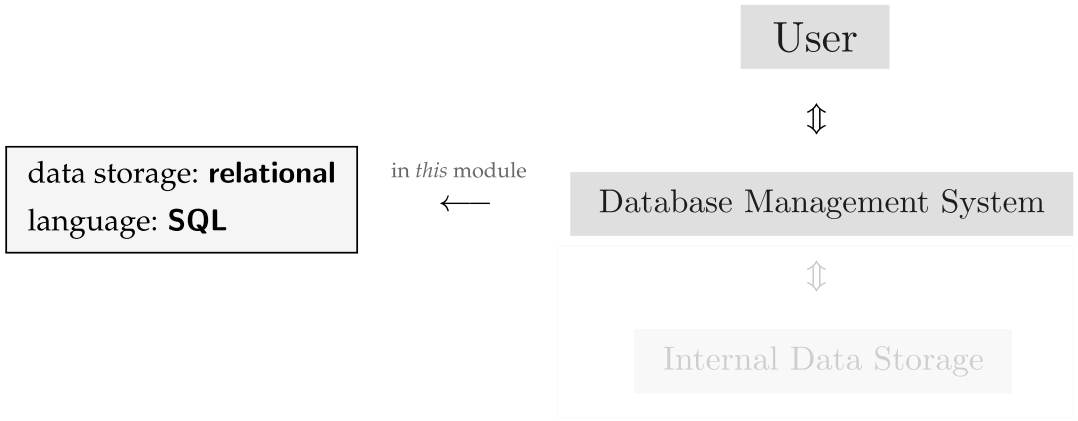
confidentiality

Databases

A database management system (DBMS) consists of:

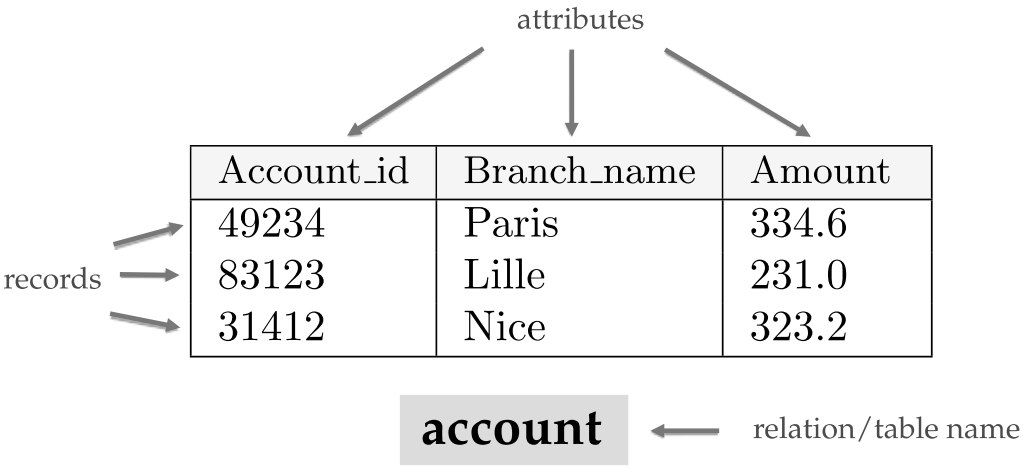
storage of data—called a **database**

language/software to manipulate/search a database



Relations (Tables)

A relation is just a table of data



Relations (Tables)

Account_id	Branch_name	Amount
49234	Paris	334.6
83123	Lille	231.0
31412	Nice	323.2

account

each row in the table represents one record

| 83123 | Lille | 231.0 |

each record is a sequence of individual data

83123	Lille	231.0

Rows

Account_id	Branch_name	Amount
49234	Paris	334.6
83123	Lille	231.0
31412	Nice	323.2

account

record

each record is a *sequence* of individual data

(Account_id: 83123, Branch_name: Lille, Amount: 231.0)

or a shorter way:

(83123, Lille, 231.0)

mathematical name : a **tuple**

Columns

account

Account_id	Branch_name	Amount
49234	Paris	334.6
83123	Lille	231.0
31412	Nice	323.2

attribute

also called its **domain**

each attribute has a data type

Account_id	integer
Branch_name	string
Amount	float

technical phrase: 'domain is atomic'

indivisible units of data

handle it as a *single piece of data*

74242	Nice, Marseille	423.2

✗

null = value unknown or unspecified

Relations (Tables)

account

Account_id	Branch_name	Amount
49234	Paris	334.6
83123	Lille	231.0
31412	Nice	323.2

relation/table

there is no order among tuples

no duplicate tuples

(Account_id: 83123, Branch_name: Lille, Amount: 231.0)
(Account_id: 49234, Branch_name: Paris, Amount: 334.6)
(Account_id: 31412, Branch_name: Nice, Amount: 323.2)

⇕ **same!**

(Account_id: 31412, Branch_name: Nice, Amount: 323.2)
(Account_id: 83123, Branch_name: Lille, Amount: 231.0)
(Account_id: 49234, Branch_name: Paris, Amount: 334.6)

mathematical name :

a relation is a **set** of tuples

Relations (Tables)

To define a relation structure, we need to specify :

the list of attribute names
domain of each attribute
} called a schema

account

Account_id	Branch_name	Amount
49234	Paris	334.6
83123	Lille	231.0
31412	Nice	323.2

schema of the table: attribute name, attribute domain

table name → account (Account_id integer, Branch_name string, Amount float)

a shorter way to write this: account (Account_id, Branch_name, Amount)

Examples

schema: account (Account_id integer, Branch_name string, Amount float)

Account_id	Branch_name	Amount
49234	Paris	334.6
83123	Lille	231.0
31412	Nice	323.2
34234	Marseille	423.12
62433	Tour	391
31412	Nice	323.2

Valid table or not ? NO

Examples

schema: account (Account_id integer, Branch_name string, Amount float)

Account_id	Branch_name	Amount
49234	Paris	334.6
83123	Lille	231.0
31412	Nice	323.2
34234	Marseille	423.12
62433	Tour	391
31412	Nice	271.1

Valid table or not ? YES

No two tuples can have same values for **all** attributes

But can have same values for **some** attributes

Examples

schema: account (Account_id integer, Branch_name string, Amount float)

Account_id	Branch_name	Amount
49234	Paris	334.6
83123	Lille	231.0
31412	Nice	323.2
34234	Marseille	423.12
62433	Tour	391
31412	Nce	323.2

Valid table or not ? YES

Examples

schema: account (Account_id integer, Branch_name string, Amount float)

Account_id	Branch_name	Amount
49234	Paris	334.6
83123	Lille	231.0
31412	Nice	323.2
34234	Marseille	423.12
62433	Tour	391
31412	Nce	323.2

?
=

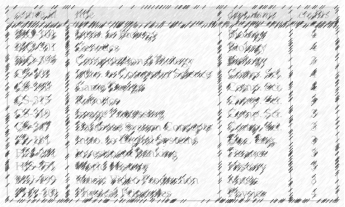
Account_id	Branch_name	Amount
49234	Paris	334.6
31412	Nce	323.2
31412	Nice	323.2
34234	Marseille	423.12
83123	Lille	231.0
62433	Tour	391

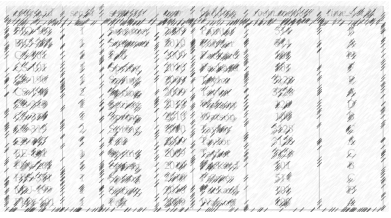
Are the two tables equal ? YES

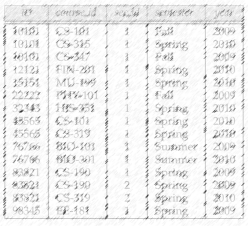
order of rows does not matter!

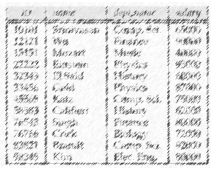
Relational Database

A relational database consists of a set of tables









dept_name	building	budget
Biology	Watson	90000
Comp. Sci.	Taylor	100000
Elec. Eng.	Taylor	85000
Finance	Painter	120000
History	Painter	50000
Music	Packard	80000
Physics	Watson	70000

We will often use the University Database as an example in this module

teacher, student, course, section, teaches, takes, department

id	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000.00
12121	Wu	Finance	90000.00
15151	Mozart	Music	40000.00
22222	Einstein	Physics	95000.00
32343	El Said	History	60000.00
33456	Gold	Physics	87000.00
45565	Katz	Comp. Sci.	75000.00
58583	Califieri	History	62000.00
76543	Singh	Finance	80000.00
76766	Crick	Biology	72000.00
83821	Brandt	Comp. Sci.	92000.00
98345	Kim	Elec. Eng.	80000.00

teacher

(from Database System Concepts)

course_id	*title*	*dept_name*	*credits*
BIO-101	Intro. to Biology	Biology	4
BIO-301	Genetics	Biology	4
BIO-399	Computational Biology	Biology	3
CS-101	Intro. to Computer Science	Comp. Sci.	4
CS-190	Game Design	Comp. Sci.	4
CS-315	Robotics	Comp. Sci.	3
CS-319	Image Processing	Comp. Sci.	3
CS-347	Database System Concepts	Comp. Sci.	3
EE-181	Intro. to Digital Systems	Elec. Eng.	3
FIN-201	Investment Banking	Finance	3
HIS-351	World History	History	3
MU-199	Music Video Production	Music	3
PHY-101	Physical Principles	Physics	4

course

(from Database System Concepts)

course_id	*sec_id*	*semester*	*year*	*building*	*room_number*	*time_slot_id*
BIO-101	1	Summer	2009	Painter	514	B
BIO-301	1	Summer	2010	Painter	514	A
CS-101	1	Fall	2009	Packard	101	H
CS-101	1	Spring	2010	Packard	101	F
CS-190	1	Spring	2009	Taylor	3128	E
CS-190	2	Spring	2009	Taylor	3128	A
CS-315	1	Spring	2010	Watson	120	D
CS-319	1	Spring	2010	Watson	100	B
CS-319	2	Spring	2010	Taylor	3128	C
CS-347	1	Fall	2009	Taylor	3128	A
EE-181	1	Spring	2009	Taylor	3128	C
FIN-201	1	Spring	2010	Packard	101	B
HIS-351	1	Spring	2010	Painter	514	C
MU-199	1	Spring	2010	Packard	101	D
PHY-101	1	Fall	2009	Watson	100	A

section

(from Database System Concepts)

id	*course_id*	*sec_id*	*semester*	*year*
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-201	1	Spring	2010
15151	MU-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010
98345	EE-181	1	Spring	2009

teaches

(from Database System Concepts)

dept_name	*building*	*budget*
Biology	Watson	90000.00
Comp. Sci.	Taylor	100000.00
Elec. Eng.	Taylor	85000.00
Finance	Painter	120000.00
History	Painter	50000.00
Music	Packard	80000.00
Physics	Watson	70000.00

department

(from Database System Concepts)

id	*name*	*dept_name*	*tot_cred*
00128	Zhang	Comp. Sci.	102
12345	Shankar	Comp. Sci.	32
19991	Brandt	History	80
23121	Chavez	Finance	110
44553	Peltier	Physics	56
45678	Levy	Physics	46
54321	Williams	Comp. Sci.	54
55739	Sanchez	Music	38
70557	Snow	Physics	0
76543	Brown	Comp. Sci.	58
76653	Aoi	Elec. Eng.	60
98765	Bourikas	Elec. Eng.	98
98988	Tanaka	Biology	120

student

(from Database System Concepts)

id	*course_id*	*sec_id*	*semester*	*year*	*grade*
00128	CS-101	1	Fall	2009	A
00128	CS-347	1	Fall	2009	A-
12345	CS-101	1	Fall	2009	C
12345	CS-190	2	Spring	2009	A
12345	CS-315	1	Spring	2010	A
12345	CS-347	1	Fall	2009	A
19991	HIS-351	1	Spring	2010	B
23121	FIN-201	1	Spring	2010	C+
44553	PHY-101	1	Fall	2009	B-
45678	CS-101	1	Fall	2009	F
45678	CS-101	1	Spring	2010	B+
45678	CS-319	1	Spring	2010	B
54321	CS-101	1	Fall	2009	A-
54321	CS-190	2	Spring	2009	B+
55739	MU-199	1	Spring	2010	A-
76543	CS-101	1	Fall	2009	A
76543	CS-319	2	Spring	2010	A
76653	EE-181	1	Spring	2009	C
98765	CS-101	1	Fall	2009	C-
98765	CS-315	1	Spring	2010	B
98988	BIO-101	1	Summer	2009	A
98988	BIO-301	1	Summer	2010	

takes

(from Database System Concepts)

course

course_id	*title*	*dept_name*	*credits*
BIO-101	Intro. to Biology	Biology	4
BIO-301	Genetics	Biology	4
BIO-399	Computational Biology	Biology	3
CS-319	Image Processing	Comp. Sci.	3
CS-347	Database System Concepts	Comp. Sci.	3
EE-181	Intro. to Digital Systems		
FIN-201	Investment Banking		
HIS-351	World History	History	3
MU-199	Music Video Production	Music	3
PHY-101	Physical Principles	Physics	4

course (course_id, title, dept_name, credits)

section

course_id	*sec_id*	*semester*	*year*	*building*	*room_number*	*time_slot_id*
BIO-101	1	Summer	2009	Painter	514	B
BIO-301	1	Summer	2010	Painter	514	A
CS-101	1	Fall	2009	Packard	101	H
CS-101	1	Spring	2010	Packard	101	F
CS-190	1	Spring	2009	Taylor	3128	E
CS-190	2	Spring	2009	Taylor	3128	A
CS-315	1	Spring	2010	Watson	120	D
CS-319	1	Spring	2010	Watson	100	B
FIN-201	1	Spring	2010	Packard	101	B
HIS-351	1	Spring	2010	Painter	514	C
MU-199	1	Spring	2010	Packard	101	D
PHY-101	1	Fall	2009	Watson	100	A

section (course_id, sec_id, semester, year, building, room_number, time_slot_id)

teaches

id	*course_id*	*sec_id*	*semester*	*year*
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-201	1	Spring	2010
15151	MU-199	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010
98345	EE-181	1	Spring	2009

teaches (id, course_id, sec_id, semester, year)

teacher

id	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000.00
12121	Wu	Finance	90000.00
15151	Mozart	Music	40000.00
22222	Einstein	Physics	95000.00
32343	El Said	History	60000.00
33456	Gold	Physics	87000.00
45565	Katz	Comp. Sci.	75000.00
58583	Califieri	History	62000.00
76543	Singh	Finance	80000.00
98345	Kim	Elec. Eng.	80000.00

teacher (id, name, dept_name, salary)

department

dept_name	*building*	*budget*
Biology	Watson	90000.00
Comp. Sci.	Taylor	100000.00
Elec. Eng.	Taylor	85000.00
Music	Packard	80000.00
Physics	Watson	70000.00

department (dept_name, building, budget)