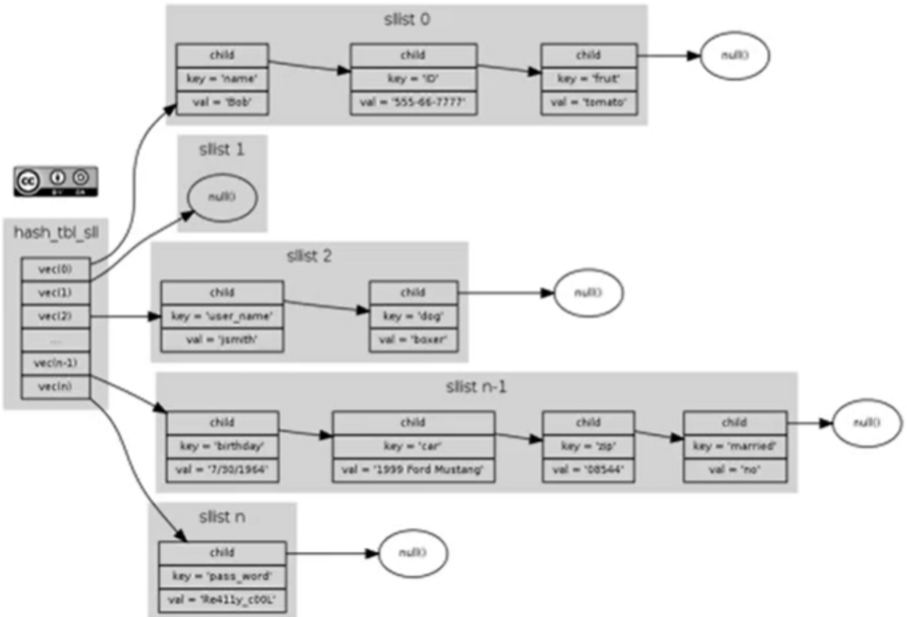
Chapter 2: Relational Model

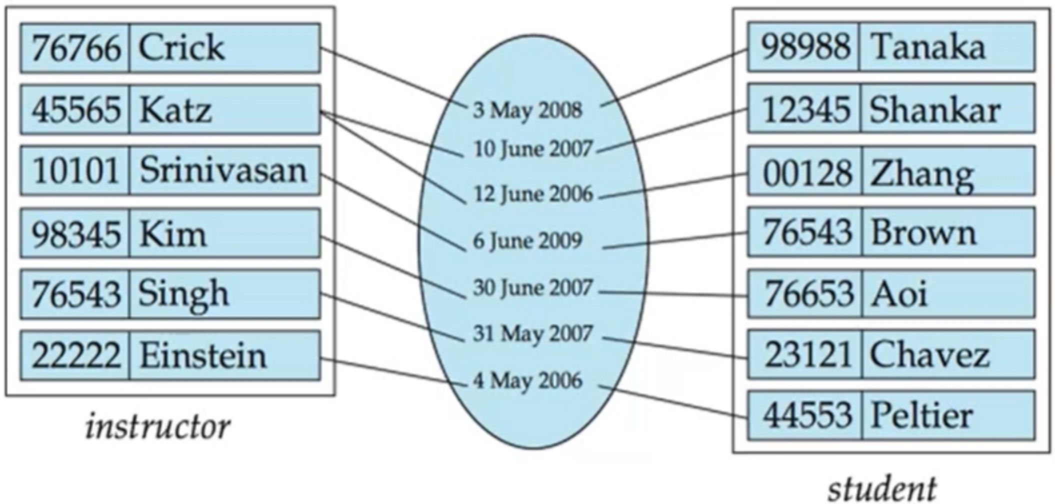
- Structure of Relational Databases
- Database Schema
- Keys
- Schema Diagrams
- Relational Query Languages
- Relational Operations

Data Model

- The framework/formalism for representing data and their relationships

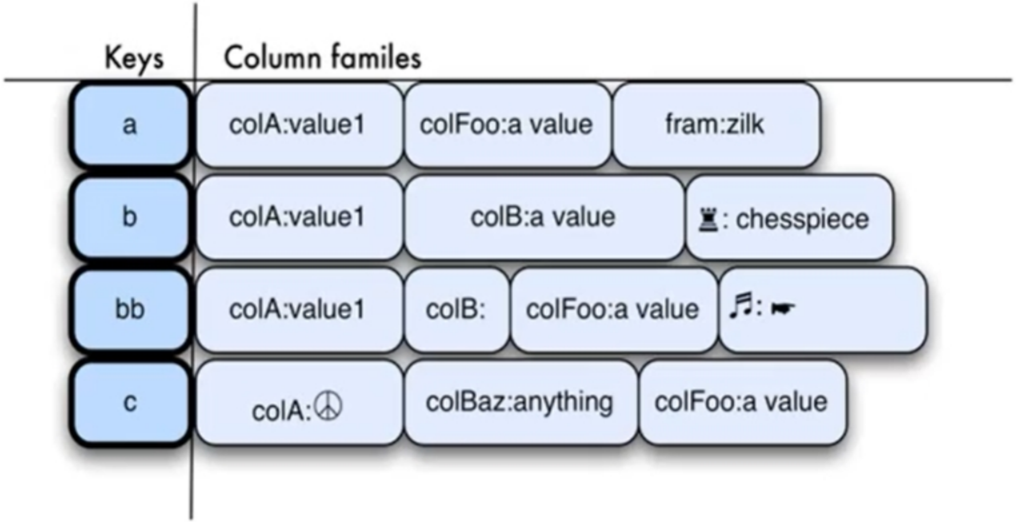


Data Model

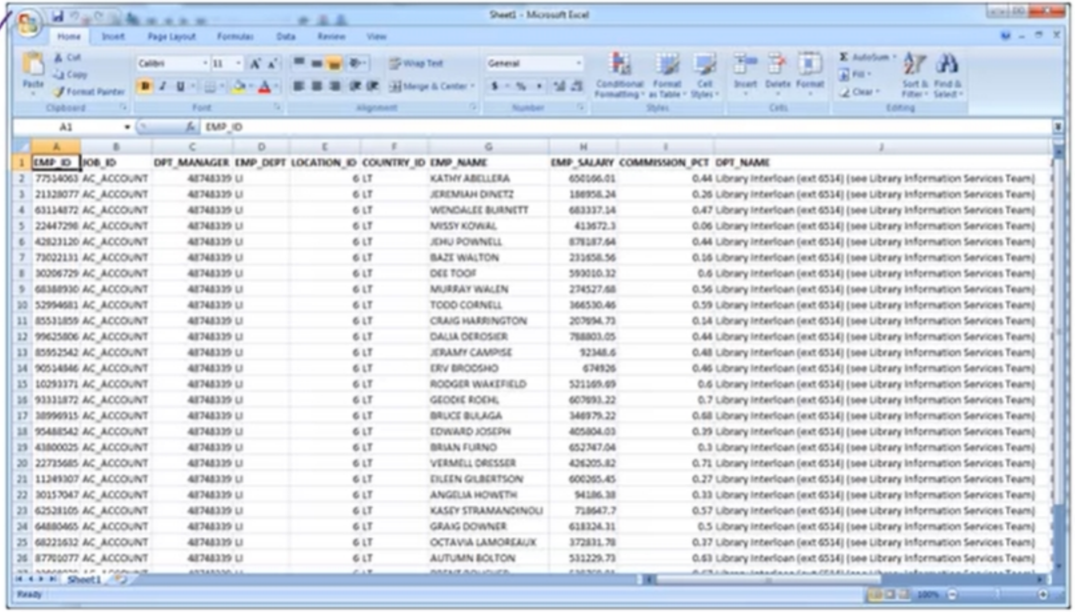


76766 Crick
45565 Katz
10101 Srinivasan
98345 Kim
76543 Singh
22222 Einstein
instructor
3 May 2008
10 June 2007
12 June 2006
6 June 2009
30 June 2007
31 May 2007
4 May 2006
98988 Tanaka
12345 Shankar
00128 Zhang
76543 Brown
76653 Aoi
23121 Chavez
44553 Peltier
student

Data Model


Keys
Column familes
a
colA:value1
colFoo:a value
fram:zilk
b
colA:value1
colB:a value
♜: chesspiece
bb
colA:value1
colB:
colFoo:a value
♬: ☛
c
colA:☮
colBaz:anything
colFoo:a value

Data Model



	EMP_ID	JOB_ID	DPT_MANAGER	EMP_DEPT	LOCATION_ID	COUNTRY_ID	EMP_NAME	EMP_SALARY	COMMISSION_PCT	DPT_NAME
2	77514063	AC_ACCOUNT	48748339	U	6	LT	KATHY ABELLERA	650166.01	0.44	Library Interloan (ext 6514) (see Library Information Services Team)
3	21328077	AC_ACCOUNT	48748339	U	6	LT	JEREMIAH DINETZ	188958.24	0.26	Library Interloan (ext 6514) (see Library Information Services Team)
4	63114872	AC_ACCOUNT	48748339	U	6	LT	WENDALEE BURNETT	683337.14	0.47	Library Interloan (ext 6514) (see Library Information Services Team)
5	22447298	AC_ACCOUNT	48748339	U	6	LT	MISSY KOWAL	413672.3	0.06	Library Interloan (ext 6514) (see Library Information Services Team)
6	42823120	AC_ACCOUNT	48748339	U	6	LT	JEHU POWNELL	878187.64	0.44	Library Interloan (ext 6514) (see Library Information Services Team)
7	73022131	AC_ACCOUNT	48748339	U	6	LT	BAZE WALTON	231658.56	0.16	Library Interloan (ext 6514) (see Library Information Services Team)
8	30206729	AC_ACCOUNT	48748339	U	6	LT	DEE TOOF	593010.32	0.6	Library Interloan (ext 6514) (see Library Information Services Team)
9	68388930	AC_ACCOUNT	48748339	U	6	LT	MURRAY WALEN	274527.68	0.56	Library Interloan (ext 6514) (see Library Information Services Team)
10	52994681	AC_ACCOUNT	48748339	U	6	LT	TODD CORNELL	366530.46	0.59	Library Interloan (ext 6514) (see Library Information Services Team)
11	85531859	AC_ACCOUNT	48748339	U	6	LT	CRAIG HARRINGTON	207894.73	0.14	Library Interloan (ext 6514) (see Library Information Services Team)
12	99625806	AC_ACCOUNT	48748339	U	6	LT	DALIA DEROSIER	788803.05	0.44	Library Interloan (ext 6514) (see Library Information Services Team)
13	85952542	AC_ACCOUNT	48748339	U	6	LT	JERAMY CAMPISE	92348.6	0.48	Library Interloan (ext 6514) (see Library Information Services Team)
14	90514846	AC_ACCOUNT	48748339	U	6	LT	ERV BRODSHO	674926	0.46	Library Interloan (ext 6514) (see Library Information Services Team)
15	10293371	AC_ACCOUNT	48748339	U	6	LT	RODGER WAKEFIELD	521169.69	0.6	Library Interloan (ext 6514) (see Library Information Services Team)
16	93331872	AC_ACCOUNT	48748339	U	6	LT	GEODIE ROEHL	607693.22	0.7	Library Interloan (ext 6514) (see Library Information Services Team)
17	38996915	AC_ACCOUNT	48748339	U	6	LT	BRUCE BULAGA	348979.22	0.68	Library Interloan (ext 6514) (see Library Information Services Team)
18	95488542	AC_ACCOUNT	48748339	U	6	LT	EDWARD JOSEPH	405804.03	0.39	Library Interloan (ext 6514) (see Library Information Services Team)
19	43800025	AC_ACCOUNT	48748339	U	6	LT	BRIAN FURNO	652747.04	0.3	Library Interloan (ext 6514) (see Library Information Services Team)
20	22735685	AC_ACCOUNT	48748339	U	6	LT	VERMELL DRESSER	426205.82	0.71	Library Interloan (ext 6514) (see Library Information Services Team)
21	11249307	AC_ACCOUNT	48748339	U	6	LT	EILEEN GILBERTSON	600265.45	0.27	Library Interloan (ext 6514) (see Library Information Services Team)
22	30157047	AC_ACCOUNT	48748339	U	6	LT	ANGELIA HOWETH	94186.38	0.33	Library Interloan (ext 6514) (see Library Information Services Team)
23	62528105	AC_ACCOUNT	48748339	U	6	LT	KASEY STRAMANDINOLI	718647.7	0.57	Library Interloan (ext 6514) (see Library Information Services Team)
24	64880465	AC_ACCOUNT	48748339	U	6	LT	GRAIG DOWNER	618324.31	0.5	Library Interloan (ext 6514) (see Library Information Services Team)
25	68221632	AC_ACCOUNT	48748339	U	6	LT	OCTAVIA LAMOREAUX	372831.78	0.37	Library Interloan (ext 6514) (see Library Information Services Team)
26	87701077	AC_ACCOUNT	48748339	U	6	LT	AUTUMN BOLTON	531229.73	0.63	Library Interloan (ext 6514) (see Library Information Services Team)

Data Model - Relational

attributes
(or columns)

ID	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

tuples
(or rows)

Relation (in Set Theory)

- **Binary relation** *R* on a set *A*
  is a collection of ordered pairs of elements of *A*
  - $A = \{ 1, 2, 3 \}$
  - $< \; = \{ <1, 2>, <1, 3>, <2, 3> \}$
  - $\leq \; = \{ <1, 2>, <1, 3>, <2, 3>, <1, 1>, <2, 2>, <3, 3> \}$

- $R \subseteq A \times A$

- *(Bipartite) Relation R2* between elements of set *B* and set *C*
  - *B* = {*미국, 일본, 중국, 한국*}, *C* = {*베이징, 서울, 워싱턴, 토교*}
  - *수도* = { < *미국, 워싱턴*>, <*일본, 토교*>, <*중국, 베이징*>, <*한국, 서울*> }

Basic Structure of a Relation

- Formally,
  given sets $D_1, D_2, \ldots D_n$ a *relation* r is a subset of $D_1 \times D_2 \times \ldots \times D_n$

  $r \subseteq D_1, \mathrm{X} \ldots\ldots \mathrm{X}\, D_n$ (n: degree of r)

  or

  $r = \{ <d_1, \ldots, d_n> \mid d_1 \in D_1, \ldots, d_n \in D_n \}$ (set of tuples)

- Example: *customer-name* = {Jones, Smith, Curry, Lindsay}
  customer-street = {Main, North, Park}
  customer-city = {Harrison, Rye, Pittsfield}

  Then r = { (Jones, Main, Harrison), (Smith, North, Rye),
  (Curry, North, Rye), (Lindsay, Park, Pittsfield)}

  is a relation over *customer-name* x *customer-street* x *customer-city*

Attribute Types

- Each attribute of a relation has a name
- The set of allowed values for each attribute is called the domain of the attribute
- Attribute values are (normally) required to be atomic,

  that is, indivisible
  - E.g. multivalued attribute values are not atomic
  - E.g. composite attribute values are not atomic

Relation Schema

- $A_1, A_2, \ldots, A_n$ are *attributes*
- $R = (A_1, A_2, \ldots, A_n)$ is a *relation schema*

  e.g.

  Customer-schema = (*customer-name, customer-street, customer-city*)

- $r(R)$ is a *relation* (*variable*) on the *relation schema R*

  e.g.

  customer(Customer-schema)

Relation Instance

- The current values (*relation instance*) of a relation are specified by a table
- An element *t* of *r* is a *tuple*,
  represented by a *row* in a table

attributes

customer-name	*customer-street*	customer-city
Jones	Main	Harrison
Smith	North	Rye
Curry	North	Rye
Lindsay	Park	Pittsfield

tuples

customer

Relations are Unordered

- Order of tuples is irrelevant
  (tuples may be stored in an arbitrary order)

- E.g. *instructor* relation with unordered tuples

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

Relational Database

- A database consists of multiple relations
- Information about an enterprise is broken up into parts, with each relation storing one part of the information

  E.g.: *instructor*: information about teachers in the university
  student: information about students
  advisor: information about which instructor advises which students

- Storing all information as a single relation is not a good idea

  univ (*instructor -ID, name, dept_name, salary, student_Id*, ..)

- Database design (Chapter 7 & 8) deals with how to decide on the relational schemas

A relational database for a university

ID	name	dept_name	salary
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
4556			
9834			
7676			
1010			
5858			
8382			
1515			
3345			
7654			

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

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 Desig		
CS-315	Robotics		
CS-319	Image Proc		
CS-347	Database Sy		
EE-181	Intro. to Dig		
FIN-201	Investment		
HIS-351	World Histo		
MU-199	Music Vide		
PHY-101	Physical Pri		

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	
CS-190	2	Spring	2009	Taylor	3128	
CS-315	1	Spring	2010	Watson	120	
CS-319	1	Spring	2010	Watson	100	
CS-319	2	Spring	2010	Taylor	3128	
CS-347	1	Fall	2009	Taylor	3128	
EE-181	1	Spring	2009	Taylor	3128	
FIN-201	1	Spring	2010	Packard	101	
HIS-351	1	Spring	2010	Painter	514	
MU-199	1	Spring	2010	Packard	101	
PHY-101	1	Fall	2009	Watson	100	

course_id	prereq_id
BIO-301	BIO-101
BIO-399	BIO-101
CS-190	CS-101
CS-315	CS-101
CS-319	CS-101
CS-347	CS-101
EE-181	PHY-101

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

Keys

- Let $K \subseteq R$
- *K* is a ***superkey*** of *R*

  if values for *K* are sufficient to identify a unique tuple of each possible relation *r(R)*.
- By "possible *r*" we mean a relation *r* that could exist in the enterprise we are modeling.
  - Example: {*customer-name*, *customer-street*} and {*customer-name*}

    are both superkeys of *Customer*, if no two customers can possibly have the same name.
- *K* is a ***candidate key*** if *K* is minimal
  - Example: {*customer-name*} is a candidate key for *Customer*, since it is a superkey, and no subset of it is a superkey.

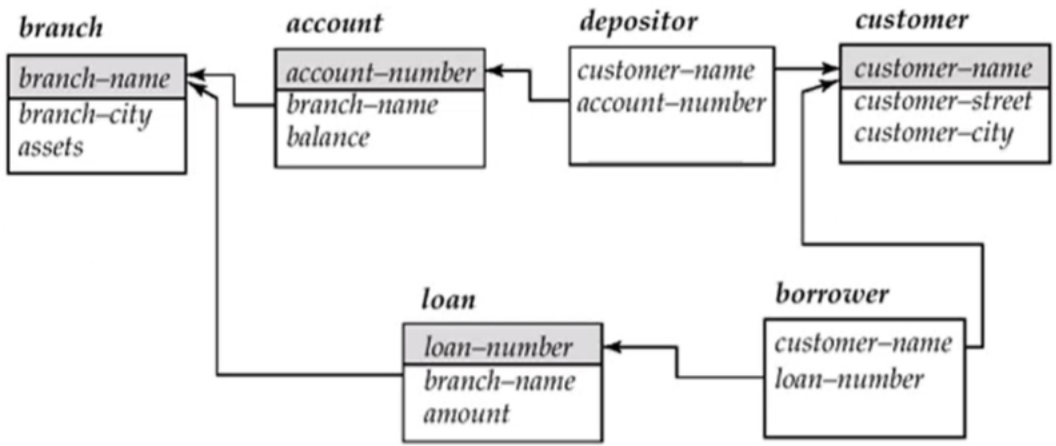
Keys

- **Primary key**: one of the candidate keys is selected to be the primary key of the table
  - which one?

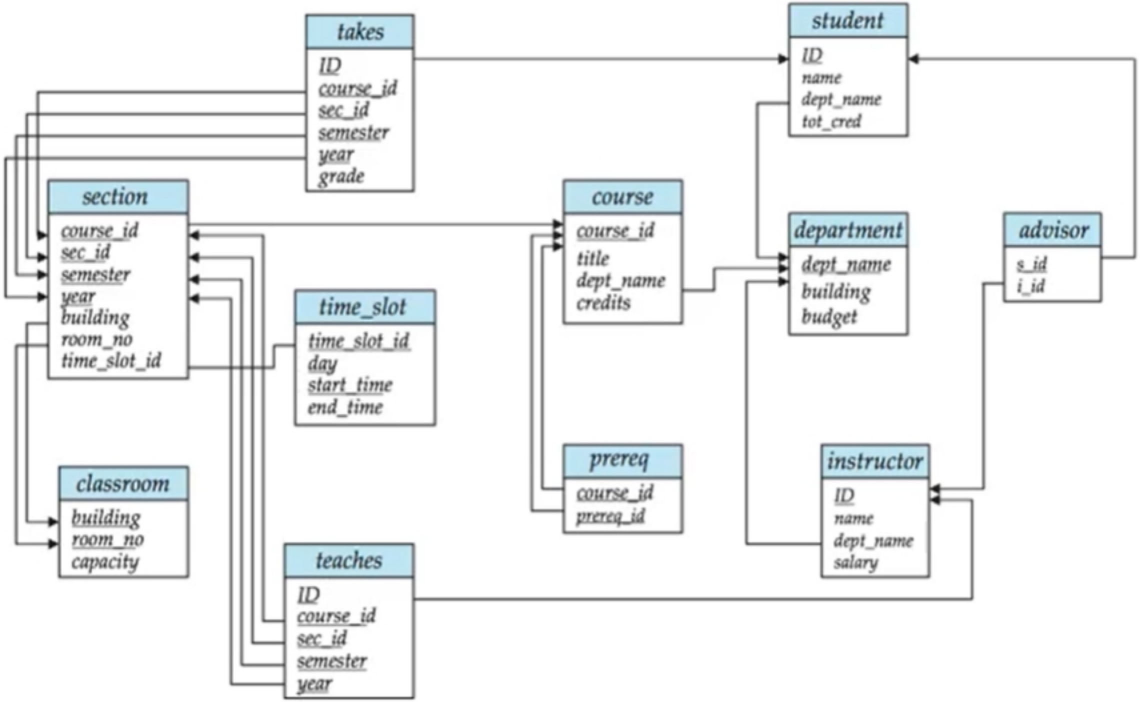
- **Foreign key**: Value in one relation must appear in another
  - **Referencing** relation
  - **Referenced** relation

course_id	*prereq_id*
BIO-301	BIO-101
BIO-399	BIO-101
CS-190	CS-101
CS-315	CS-101
CS-319	CS-101
CS-347	CS-101
EE-181	PHY-101

Schema Diagram for a Banking Enterprise



Schema Diagram for the University



Query Languages

- Language in which user requests information from the database.
  - procedural
    - specify what data are needed and how to get those data
  - nonprocedural
    - declarative
    - specify only what data are needed
- "Pure" languages:
  - Relational Algebra
  - Tuple Relational Calculus
  - Domain Relational Calculus
- Pure languages form underlying basis of query languages that people use.

Relational Algebra

- Algebra : operators and operands
  - Relational algebra
    - operands : relations
    - operators : basic operators (+ additional operations)
      - take two or more relations as inputs and give a new relation as a result.
- Procedural language
- Operators

select	join
project	division
union	assignment
set difference	…
Cartesian product	
rename	

Examples of Relational Operators

Symbol (Name)	Example of Use
σ (Selection)	$\sigma_{\text{salary>=85000}}(\textit{instructor})$
	Return rows of the input relation that satisfy the predicate.
Π (Projection)	$\Pi_{\textit{ID, salary}}(\textit{instructor})$
	Output specified attributes from all rows of the input relation. Remove duplicate tuples from the output.
$\bowtie$ (Natural Join)	$\textit{instructor} \bowtie \textit{department}$
	Output pairs of rows from the two input relations that have the same value on all attributes that have the same name.
$\times$ (Cartesian Product)	$\textit{instructor} \times \textit{department}$
	Output all pairs of rows from the two input relations (regardless of whether or not they have the same values on common attributes)
$\cup$ (Union)	$\Pi_{\textit{name}}(\textit{instructor}) \cup \Pi_{\textit{name}}(\textit{student})$
	Output the union of tuples from the two input relations.

Select Operation – selection of tuples

- Relation *r* :

A	*B*	*C*	*D*
α	α	1	7
α	β	5	7
β	β	12	3
β	β	23	10

- $\sigma_{A=B \wedge D>5}(r)$:

A	*B*	*C*	*D*
α	α	1	7
β	β	23	10

Project Operation – selection of columns

- Relation *r*:

A	B	C
α	10	1
α	20	1
β	30	1
β	40	2

- $\Pi_{A,C}(r)$

A	C
α	1
α	1
β	1
β	2

=

A	C
α	1
β	1
β	2

Union Operation – merging two relations

- Relations *r, s:*

A	B
α	1
α	2
β	1

r

A	B
α	2
β	3

s

- $r \cup s$:

A	B
α	1
α	2
β	1
β	3

Set Difference Operation

- Relations *r, s:*

A	B
α	1
α	2
β	1

r

A	B
α	2
β	3

s

- *r – s:*

A	B
α	1
β	1

Set-Intersection Operation

- Relation *r*, *s*:

r

A	B
α	1
α	2
β	1

s

A	B
α	2
β	3

- $r \cap s$

A	B
α	2

Joining two relations: Cartesian-Product Op.

- Relations *r, s:*

r

A	B
α	1
β	2

s

C	D	E
α	10	a
β	19	a
β	20	b
γ	10	b

- *r* x *s:*

A	B	C	D	E
α	1	α	10	a
α	1	β	19	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	19	a
β	2	β	20	b
β	2	γ	10	b

Joining two relations: Natural-Join Operation

- Relations r, s:

r

A	B	C	D
α	1	α	a
β	2	γ	a
γ	4	β	b
α	1	γ	a
δ	2	β	b

s

B	D	E
1	a	α
3	a	β
1	a	γ
2	b	δ
3	b	ϵ

- $r \bowtie s$

A	B	C	D	E
α	1	α	a	α
α	1	α	a	γ
α	1	γ	a	α
α	1	γ	a	γ
δ	2	β	b	δ

Composition of Operations

- Can build expressions using multiple operations
- Example: $\sigma_{A=C}(r \times s)$
- $r \times s$

r

A	B
α	1
β	2

s

C	D	E
α	10	a
β	19	a
β	20	b
γ	10	b

A	B	C	D	E
α	1	α	10	a
α	1	β	19	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	19	a
β	2	β	20	b
β	2	γ	10	b

- $\sigma_{A=C}(r \times s)$

A	B	C	D	E
α	1	α	10	a
β	2	β	19	a
β	2	β	20	b

END OF CHAPTER 2