What would be the proper Candidate Key/Primary Key?

- F19Courses (CID, Title, Time, Location, Cap)

Subject	Catalog #	Units	Title
CS	115	0.5	Intro to Computer Science 1

Notes: Choose LAB section for Related 1.

Class	Comp	Sec	Camp Loc	Assoc Class	Rel 1	Rel 2	Enrl Cap	Enrl Tot	Wait Cap	Wait Tot	Time	Days/Date	Bldg Room	Instructor
6119	LEC	001	UW U	1		201	90	84	0	0	08:30-09:50	TTh	PHY 235	Case,Lori
Reserve: Year 1 students							60	50						
6120	LEC	002	UW U	2		201	90	63	0	0	04:00-05:20	TTh	MC 2035	Hideg,Valentina
Reserve: Year 1 students							60	46						
6121	LEC	003	UW U	3		201	90	88	0	0	10:00-11:20	TTh	PHY 235	Case,Lori
Reserve: Year 1 students							60	53						
6122	LEC	004	UW U	4		201	90	88	0	0	10:00-11:20	TTh	MC 4020	Vasiga,Troy Michael
Reserve: Year 1 students							60	54						
6358	LEC	005	UW U	5		201	90	60	0	0	11:30-12:50	TTh	MC 2017	Graham,Sandy
Reserve: Yr 1 Honours Geomatics							20	9						
Reserve: Year 1 students							50	22						
6369	LEC	006	UW U	6		201	90	55	0	0	10:00-11:20	TTh	MC 2017	Akinyemi,John
Reserve: Year 1 students							60	37						

- F20Courses (CID, Title, Time, Location, Cap)

Subject	Catalog #	Units	Title
CS	115	0.5	Intro to Computer Science 1

Notes: Choose LAB section for Related 1.

Class	Comp	Sec	Camp Loc	Assoc Class	Rel 1	Rel 2	Enrl Cap	Enrl Tot	Wait Cap	Wait Tot	Time	Days/Date	Bldg Room	Instructor
6297	LEC	041	ONLN ONLINE	41		241	150	149	0	0	TBA			Kaplan,Craig
Reserve: Year 1 students							90	89						
6298	LEC	042	ONLN ONLINE	42		241	150	117	0	0	TBA			Kaplan,Craig
Reserve: Year 1 students							90	78						
6299	LEC	043	ONLN ONLINE	43		241	150	106	0	0	TBA			Pretti,J.P.
Reserve: Year 1 students							90	78						
6300	LEC	044	ONLN ONLINE	44		241	150	108	0	0	TBA			Pretti,J.P.
Reserve: Year 1 students							90	81						
6525	LEC	045	ONLN ONLINE	45		241	150	118	0	0	TBA			Vasiga,Troy Michael
Reserve: Yr 1 Honours Geomatics							20	14						
Reserve: Year 1 students							90	87						

Review of Lecture 2-1

- Relational database: tables
  - Set, simple attribute
  - Order of row vs. order of column
- ICs
  - Domain
  - PK
  - FK
    - Used to associate table with table
    - Referencing table and referenced table

Lecture 2-2: Today's Plan

- Introduction to Relational Algebra
  - Select
  - Project
  - Assignment
  - Union
  - Intersect
  - Minus
- Exercise

RELATIONAL ALGEBRA AND CALCULUS

CHAPTER 03

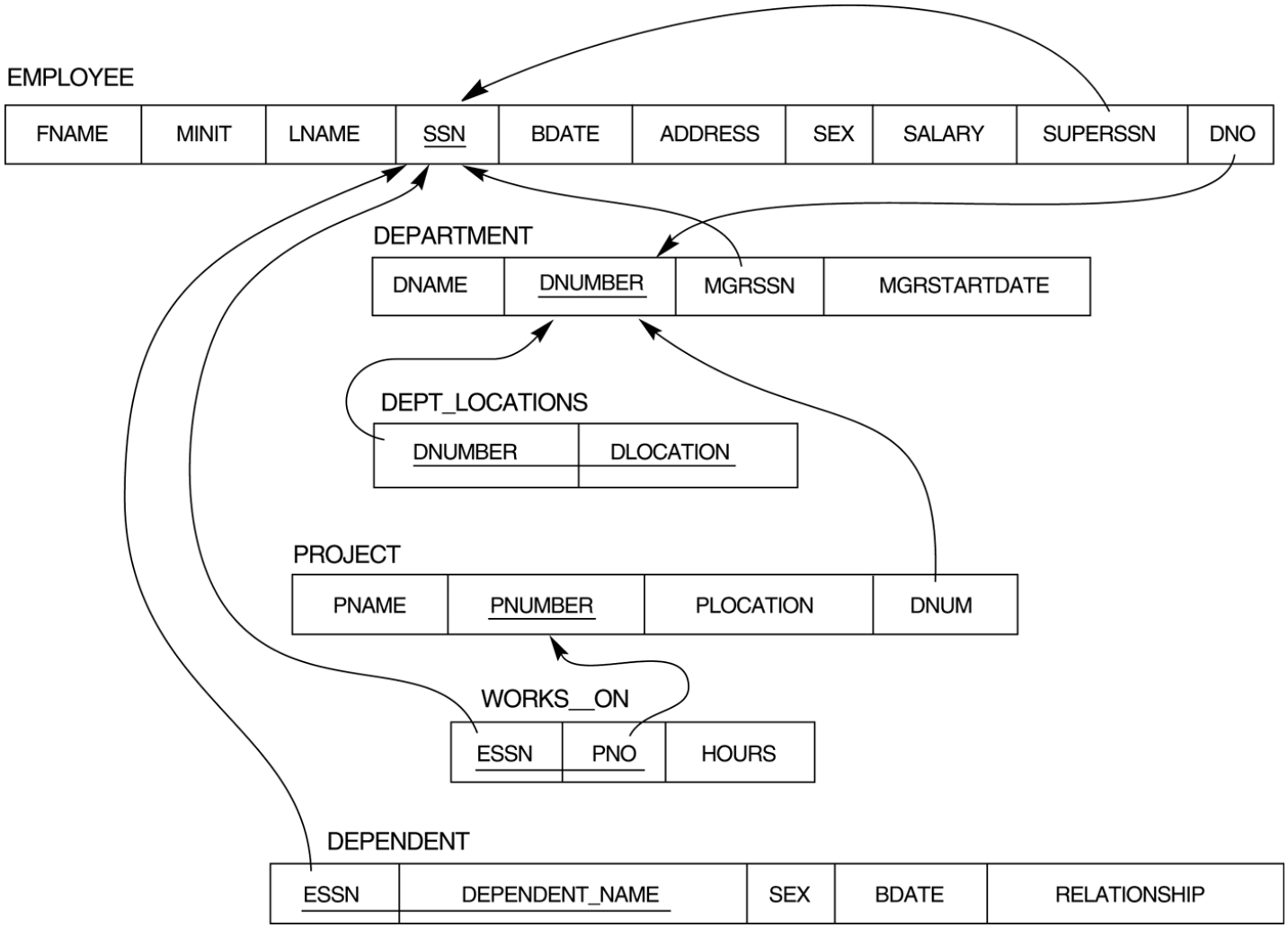
Collin Roberts

University of Waterloo

Recommended Readings

- Textbook Chapter 06

Company Relational Schema



Union
Join
Natural
Exercise
Review
Optimize

Populated Database-- Fig.5.6

EMPLOYEE	FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
	John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
	Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
	Alicia	J	Zelaya	999887777	1968-07-19	3321 Castle, Spring, TX	F	25000	987654321	4
	Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
	Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
	Joyce	A	English	453453453	1972-07-31	5631 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	55000	null	1

DEPT_LOCATIONS	DNUMBER	DLOCATION
	1	Houston
	4	Stafford
	5	Bellaire
	5	Sugarland
	5	Houston

DEPARTMENT	DNAME	DNUMBER	MGRSSN	MGRSTARTDATE
	Research	5	333445555	1988-05-22
	Administration	4	987654321	1995-01-01
	Headquarters	1	888665555	1981-06-19

WORKS_ON	ESSN	PNO	HOURS
	123456789	1	32.5
	123456789	2	7.5
	666884444	3	40.0
	453453453	1	20.0
	453453453	2	20.0
	333445555	2	10.0
	333445555	3	10.0
	333445555	10	10.0
	333445555	20	10.0
	999887777	30	30.0
	999887777	10	10.0
	987987987	10	35.0
	987987987	30	5.0
	987654321	30	20.0
	987654321	20	15.0
	888665555	20	null

PROJECT	PNAME	PNUMBER	PLOCATION	DNUM
	ProductX	1	Bellaire	5
	ProductY	2	Sugarland	5
	ProductZ	3	Houston	5
	Computerization	10	Stafford	4
	Reorganization	20	Houston	1
	Newbenefits	30	Stafford	4

DEPENDENT	ESSN	DEPENDENT_NAME	SEX	BDATE	RELATIONSHIP
	333445555	Alice	F	1986-04-05	DAUGHTER
	333445555	Theodore	M	1983-10-25	SON
	333445555	Joy	F	1958-05-03	SPOUSE
	987654321	Abner	M	1942-02-28	SPOUSE
	123456789	Michael	M	1988-01-04	SON
	123456789	Alice	F	1988-12-30	DAUGHTER
	123456789	Elizabeth	F	1967-05-05	SPOUSE

Relational Query Language

- A major strength of the relational model: supports simple, powerful *querying* of data.
- Queries can be written intuitively, and the DBMS is responsible for efficient evaluation.
  - The key: precise semantics for relational queries.
  - Allows the optimizer to extensively re-order operations, and still ensure that the answer does not change.

Procedural vs. Non-procedural

- Two major categories
  - Procedural --- focus on:
    - A sequence of access operations is specified
  - Non-procedural --- focus on:
    - What is to be retrieved, rather than how it is to be retrieved
    - *Example*: "all customers in Provo with zero account balances"
- Examples of each category
  - Procedural: relational algebra
    - Relations computed by evaluation of infix expressions
  - Non-procedural: relational calculus
    - Relations determined by logical specifications

Relational Query Language

- Relational Algebra: procedural, focus on a sequence of actions to access the data
- Relational Calculus: declarative, let users specify what they want rather than how to compute it

Query language is NOT a programming language!

What is Algebra?

- Mathematical system consisting of:
  - *Operands* --- variables or values from which new values can be constructed.
  - *Operators* --- symbols denoting procedures that construct new values from given values.

Relational Algebra

- Use to specify retrieval request
  - Operand: Relation treated as set
  - Operator: Set operators
- The result of a retrieval is a new relation
  - Each relational operator takes in relations and produces a new relation
- A sequence of relational algebra operations forms a relational algebra expression

8 – 9 / (3 * 2)

- Part of the model for SQL
  - Explain semantics formally
  - Basis for implementations
  - Fundamental to query optimization

Relational Algebra Operators

- Basic operations over relations:
  - *Selection* (σ) Selects a subset of rows from relation.
  - *Projection* (π) Deletes unwanted columns from relation.
  - *Cross-product* ($\times$) Allows us to combine two relations.
  - *Set-difference* (—) Tuples in reln. 1, but not in reln. 2.
  - *Union* ($\cup$) Tuples in reln. 1 and in reln. 2.
- Additional operations:
  - Intersection, *join*, division, renaming: Not essential, but (very!) useful.
- Since each operation returns a relation, operations can be *composed*! (Algebra is "closed".)

How would you do this manually?

- Find employees who work in department #4?

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 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	55000	NULL	1

DEPARTMENT

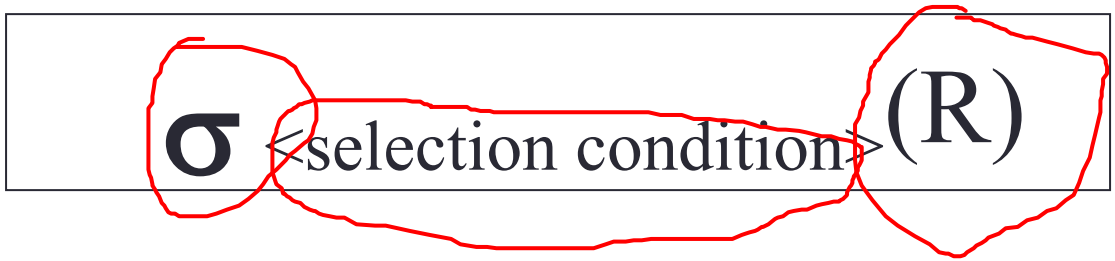
Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

Select Operation

- A unary relational operations
- Used to select a *subset* of the tuples from a relation that satisfy a **selection condition**.
  - Boolean expression



Boolean Condition

- A statement that can be true or false depending on the input
- Typical form: some value compare with other value (>, <, =, !=, >=, <=)
  - Example: Balance > 1000
- Can be combined using AND, OR, NOT
  - Example: (Balance >= 1000) AND (Owner != ‘J Smith’)
- Other forms will be introduced later in SQL

Select: Example

σ = select

Sample query: $\sigma_{\text{Type = "savings"}}$ Account

Account	Number	Owner	Balance	Type
	101	J. Smith	1000.00	checking
	102	W. Wei	2000.00	checking
	103	J. Smith	5000.00	savings
	104	M. Jones	1000.00	checking
	105	H. Martin	10,000.00	checking

	Number	Owner	Balance	Type
	103	J. Smith	5000.00	savings

Select: Another Example

$\sigma_{\text{Balance} < 4000}$ Account

Account	Number	Owner	Balance	Type
	101	J. Smith	1000.00	checking
	102	W. Wei	2000.00	checking
	103	J. Smith	5000.00	savings
	104	M. Jones	1000.00	checking
	105	H. Martin	10,000.00	checking

	Number	Owner	Balance	Type
	101	J. Smith	1000.00	checking
	102	W. Wei	2000.00	checking
	104	M. Jones	1000.00	checking

Select Operation Properties

- Produces a relation S that has the same schema as R
- **Commutative:**
  $\sigma_{<condition1>}(\sigma_{<condition2>}(R)) = \sigma_{<condition2>}(\sigma_{<condition1>}(R))$
- A cascaded SELECT operation **may be applied in any order:**
  $\sigma_{<condition1>}(\sigma_{<condition2>}(\sigma_{<condition3>}(R))$
  $= \sigma_{<condition2>}(\sigma_{<condition3>}(\sigma_{<condition1>}(R)))$
- Compound condition:
  $\sigma_{<condition1>}(\sigma_{<condition2>}(\sigma_{<condition3>}(R))$
  $= \sigma_{<condition1> \text{ AND } <condition2> \text{ AND } <condition3>}(R)$
- Operation composition: Output of SELECT can be fed to another operation

How would you do this manually?

- List the last name and first name of all employees.

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 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	55000	NULL	1

DEPARTMENT

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

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

Project Operation

- A unary relational operations
- This operation selects certain *columns* from the table and discards the other columns.

$$\pi_{<\text{attribute list}>}(R)$$

- Duplicate?
  - Eliminate in RA (based on set)
  - Not so in practice (when using SQL)

Project: Example

π_{Owner} Account

Account	Number	Owner	Balance	Type
	101	J. Smith	1000.00	checking
	102	W. Wei	2000.00	checking
	103	J. Smith	5000.00	savings
	104	M. Jones	1000.00	checking
	105	H. Martin	10,000.00	checking

	Owner
	J. Smith
	W. Wei
	M. Jones
	H. Martin

Note: Projection operator eliminates *duplicates*, *Why???*

In a DBMS products, do you think duplicates should be eliminated for every query? Are they?

Example

Figure 7.8 Results of SELECT and PROJECT operations.

(a) $\sigma_{(DNO=4\ AND\ SALARY>25000)\ OR\ (DNO=5\ AND\ SALARY>30000)}(EMPLOYEE)$.

(b) $\pi_{LNAME,\ FNAME,\ SALARY}(EMPLOYEE)$. (c) $\pi_{SEX,\ SALARY}(EMPLOYEE)$

(a)

FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
Franklin	T	Wong	333445555	1955-12-08	638 Voss,Houston,TX	M	40000	888665555	5
Jennifer		Wallace	987654321	1941-06-20	291 Berry,Bellaire,TX	F	43000	888665555	4
Ramesh		Narayan	666884444	1962-09-15	975 FireOak,Humble,TX	M	38000	333445555	5

(b)

LNAME	FNAME	SALARY
Smith	John	30000
Wong	Franklin	40000
Zelaya	Alicia	25000
Wallace	Jennifer	43000
Narayan	Ramesh	38000
English	Joyce	25000
Jabbar	Ahmad	25000
Borg	James	55000

(c)

SEX	SALARY
M	30000
M	40000
F	25000
F	43000
M	38000
M	25000
M	55000

Assignment Operation - Motivation

Example: To retrieve the first name, last name, and salary of all employees who work in department number 5, we must apply a select and a project operation. We can write a single relational algebra expression as follows:

$$\pi_{\text{FNAME, LNAME, SALARY}}(\sigma_{\text{DNO=5}}(\text{EMPLOYEE}))$$

Assignment Operation

- The assignment operation (←) provides a convenient way to express complex queries.
  - Write query as a sequential program consisting of
    - a series of assignments
    - followed by an expression whose value is displayed as a result of the query.
  - Assignment must always be made to a temporary relation variable.
  - The result to the right of the ← is assigned to the relation variable on the left of the ←.
  - May use variable in subsequent expressions

Union Operation

- **Example:** To retrieve the SSNs of all employees who either work in department 5 or directly supervise an employee who works in department 5, we can use the union operation as follows:

$$R1 \leftarrow \sigma_{DNO=5}(EMPLOYEE)$$

$$RESULT1 \leftarrow \pi_{SSN}(R1)$$

$$RESULT2 \leftarrow \pi_{SUPERSSN}(R1)$$

$$RESULT \leftarrow RESULT1 \cup RESULT2$$

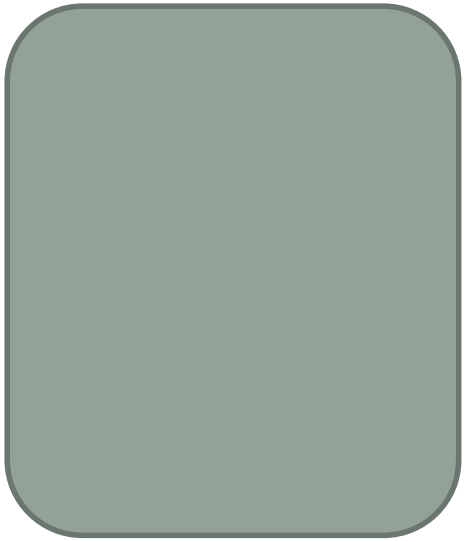
Company

R1	Fname	Mini!	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
	John	B	Smith	123 456789	1965-01-09	73 1 Fondren, Houston, TX	M	30000	333445555	5
	Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
	Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
	Joyce	A	English	453453453	1972-07-31	563 1 Rice, Houston, TX	F	25000	333445555	5

Union: Example

STUDENT	FN	LN
	Susan	Yao
	Ramesh	Shah
	Johnny	Kohler
	Barbara	Jones
	Amy	Ford
	Jimmy	Wang
	Ernest	Gilbert

INSTRUCTOR	FNAME	LNAME
	John	Smith
	Ricardo	Browne
	Susan	Yao
	Francis	Johnson
	Ramesh	Shah



STUDENT ∪ INSTRUCTOR

Union: Requirement

- What if we do a union between department and Employee:

Department ∪ Employee

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 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	55000	NULL	1

DEPARTMENT

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

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

Union Compatible

- Two relations are *union-compatible* if they have the same degree (i.e., the same number of attributes) and the corresponding attributes are defined on the same domains.

- Suppose we have these tables:

Checking-Account (c-num, c-owner, c-balance)

Savings-Account (s-num, s-owner, s-balance)

These are union-compatible tables.

- *Union, intersection, & difference require union-compatible tables*

Union Compatible: Example

∪ = union

Checking-account ∪ Savings-account

Checking-account	c-num	c-owner	c-balance
	101	J. Smith	1000.00
	102	W. Wei	2000.00
	104	M. Jones	1000.00
	105	H. Martin	10,000.00

Savings-account	s-num	s-owner	s-balance
	103	J. Smith	5000.00

	c-num	c-owner	c-balance
	101	J. Smith	1000.00
	102	W. Wei	2000.00
	104	M. Jones	1000.00
	105	H. Martin	10,000.00
	103	J. Smith	5000.00

Intersection Operation

- **Example:** Find those who are both students and instructors.

INSTRUCTOR	FNAME	LNAME
	John	Smith
	Ricardo	Browne
	Susan	Yao
	Francis	Johnson
	Ramesh	Shah

STUDENT	FN	LN
	Susan	Yao
	Ramesh	Shah
	Johnny	Kohler
	Barbara	Jones
	Amy	Ford
	Jimmy	Wang
	Ernest	Gilbert

Exercise: Intersection

Checking-account ∩ Savings-account

What's the answer to this query?

Checking-account

c-num	c-owner	c-balance
101	J. Smith	1000.00
102	W. Wei	2000.00
104	M. Jones	1000.00
105	H. Martin	10,000.00

Savings-account

s-num	s-owner	s-balance
103	J. Smith	5000.00

Difference (or Minus) Operation

Example: The figure shows the names of students who are not instructors.

INSTRUCTOR	FNAME	LNAME
	John	Smith
	Ricardo	Browne
	Susan	Yao
	Francis	Johnson
	Ramesh	Shah

STUDENT	FN	LN
	Susan	Yao
	Ramesh	Shah
	Johnny	Kohler
	Barbara	Jones
	Amy	Ford
	Jimmy	Wang
	Ernest	Gilbert

Properties of Set Operators

- Notice that both union and intersection are *commutative operations;* that is

  $R \cup S = S \cup R$, and $R \cap S = S \cap R$

- Both union and intersection can be treated as n-ary operations applicable to any number of relations as both are *associative operations;* that is

  $R \cup (S \cup T) = (R \cup S) \cup T$, and $(R \cap S) \cap T = R \cap (S \cap T)$

- The minus operation is *not commutative;* that is, in general

  $R - S \neq S - R$

SI Example

In-class Exercise #1

1. Find all the owners who have only checking account.
2. Find all the owners who have only saving account.
3. Find all the owners who have both accounts.
4. Find all the owners who have an account with us. (maybe ambiguous?)
5. Find all the owners who have some account with us.

Checking-account

c-num	c-owner	c-balance
101	J. Smith	1000.00
102	W. Wei	2000.00
104	M. Jones	1000.00
105	H. Martin	10,000.00

Savings-account

s-num	s-owner	s-balance
103	J. Smith	5000.00

Assumption: owner name is unique

Summary of Lecture 2-2

- Some basic RA operations
  - Select, Project, Assignment, Union, Intersect, Minus
- Make sure you understand the schema of the result for each operation