The Relational Algebra & Relational Calculus

政治大學
資訊科學系
沈錳坤

Content

- Relational Algebra
- Relational Calculus
  - Tuple Relational Calculus
  - Domain Relational Calculus

Relational Algebra

- Relational algebra
  - A basic set of relational model operations to enable users to specify basic retrieval requests
  - Algebra operations produce new relations
- Importance of relational algebra
  - formal foundation for relational model operations
  - a basis for implementing and optimization queries in relational DBMS
  - some of its concept are incorporated into SQL

Relational Algebra (cont.)

- Relational algebra
  - Two groups of operations
    - Set operations from mathematical set theory
      - Union
      - Intersection
      - Set Difference
      - Cartesian Product
    - Specially for RDB Operations
      - Select 'WHERE' in SQL
      - Project 'SELECT' in SQL
      - Join

希臘字母 (Greek Alphabets)

大寫	小寫	讀音	大寫	小寫	讀音	大寫	小寫	讀音
A	α	alpha	I	ι	iota	P	ρ	rho
B	β	beta	K	κ	kappa	Σ	σ , ς	sigma
Γ	γ	gamma	Λ	λ	lambda	T	τ	tau
Δ	δ	delta	M	μ	mu	Υ	υ	upsilon
E	ε	epsilon	N	ν	nu	Φ	φ	phi
Z	ζ	zeta	Ξ	ξ	xi	X	χ	khi
H	η	eta	O	o	omicron	Ψ	ϕ	psi
Θ	θ	theta	Π	π	pi	Ω	ω	omega

The Select Operation

- Select
  - Select a subset of the tuples from a relation that satisfy a selection condition
  - A filter to keep only those tuples that satisfy a qualifying condition
  - $\sigma_{\text{<selection condition>}}(R)$,
    - Selection condition
      - <attribute name> <comparison op> <constant value>
      - <attribute name> <comparison op> <attribute name>
  - e.g.
    - $\sigma_{DNO = 4}(\text{Employee})$,
    - $\sigma_{Salary>30000}(\text{Employee})$,
    - $\sigma_{\text{<DNO=4 and Salary >25000) or (DNO=5 and Salary > 30000>}}(\text{Employee})$

The Select Operation

- Property of Selection operation
  - Degree of relation resulting from a Select R operation is the same as that of R (# of columns are identical)
  - $|\sigma_C(R)| \leq |R|$ (# of rows reduces)
  - Commutative: $\sigma_{<C1>}(\sigma_{<C2>}(R)) = \sigma_{<C2>}(\sigma_{<C1>}(R))$
  - A cascade of select operation can be combined to a single select operation with a conjunctive (and) condition:

    $\sigma_{<C1>}(\sigma_{<C2>}(\ldots(\sigma_{<Cn>}(R))\ldots)) = \sigma_{<C1> \text{ and } <C2> \ldots \text{and } <Cn>}(R)$

- σ <DNO=4 and Salary >25000) or (DNO=5 and Salary > 30000>(Employee)

FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
John		Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin		Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia		Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer		Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh		Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce		English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad		Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James		Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	null	1

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

The Project Operation

- Project
  - Select certain columns from the table and discards the other columns
  - π <attribute list>(R),
    - Attribute list: a list of attributes from those of relation R
  - e.g.: $\pi_{Lname, Fname, Salary}(Employee)$
  - Duplicate elimination
    - If the attribute list includes only nonkey attributes of R, the Project operation removes any duplicate tuples
    - e.g. $\pi_{Sex, Salary}(Employee)$
  - $|\pi_L(R)| \leq |R|$ (∵ duplicate key removal)
  - $\pi_{<L1>}(\pi_{<L2>}(R)) = \pi_{<L1>}(R)$, where $L2 \supseteq L1$
  - $\pi_{<L1>}(\pi_{<L2>}(R)) \neq \pi_{<L2>}(\pi_{<L1>}(R))$

(non-commutative, valid only when L2 ⊇ L1)

- $\pi_{Lname,\ Fanme,\ Salary}(Employee)$

FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
John		Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin		Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia		Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer		Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh		Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce		English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad		Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James		Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	null	1

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

- $\pi_{Sex,\ Salary}(Employee)$

FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
John		Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin		Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia		Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer		Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh		Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce		English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad		Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James		Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	null	1

(c)

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

Sequence of Operations & Rename

- Sequence of operations
  - $\pi_{\text{Fname, Name, salary}}(\sigma_{\text{DNO=4}}(\text{Employee}))$ (project, select)

    $= \text{DEPT4\_EMPS} \leftarrow \sigma_{\text{DNO=4}}(\text{Employee})$

    $R \leftarrow \pi_{\text{Fname, Name, Salary}}(\text{DEPT4\_EMPS})$
- Rename relation name & attribute name
  - $\text{Dept4\_Emps} \leftarrow \sigma_{\text{DNO=4}}(\text{EMPLOYEE})$

    $R(\text{FirstName, LastName, Salary}) \leftarrow \pi_{\text{fname,Lname,Salary}}(\text{Dept4\_Emps})$

Set Theoretic Operations

- Binary operations from mathematical set theory
  - UNION: R1 ∪ R2
  - INTERSECTION: R1 ∩ R2
  - SET DIFFERENCE: R1 - R2
  - CARTESIAN PRODUCT: R1 X R2
- Property of ∪, ∩, -
  - Union compatible: operand relations

    R1(A1, ..., An) & R2(B1, ..., Bn)
    - degree(R1)=degree(R2)=n
    - dom(Ai)=dom(Bi) for i=1, 2, ..., n.
  - The resulting relation has the same attribute names as the *first* operand relation R1 (by convention)

Union

1st table

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

FN	LN
Susan	Yao
Ramesh	Shah
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert
John	Smith
Ricardo	Browne
Francis	Johnson

① deg(STUDENT) = deg(INSTRUCTOR) = 2

② dom(FN) = dom(FNAME)
dom(LN) = dom(LNAME)

Intersection

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

FN	LN
Susan	Yao
Ramesh	Shah

Difference (Student-Instructor)

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

FN	LN
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert

Difference (Instructor-Student)

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

FNAME	LNAME
John	Smith
Ricardo	Browne
Francis	Johnson

Set Theoretic Operations (cont.)

- Union, Intersection: commutative & associative
  - $R \cup S = S \cup R$, $R \cap S = S \cap R$
  - $R \cup (S \cup T) = (R \cup S) \cup T$
  - $R \cap (S \cap T) = (R \cap S) \cap T$
- Difference: not commutative
  - $R - S \neq S - R$

Cartesian Product

- CARTESIAN PRODUCT :
  - *R*(A1, A2, ..., Am) X *S* (B1, B2, ..., Bn) =

    T(A1, A2, ..., Am, B1, B2, ..., Bn)

    * *R*, *S* union compatible is not necessary
  - A tuple *t* exists in *T* for each combination of tuples *r* from *R* and *s* from *S* such that: *R*[A1, A2, ..., Am]=*r* and *S*[B1, B2, ..., Bn]=*s*
  - *R* has n_R tuples and *S* has n_S tuples, then T will have $n_R * n_S$ tuples.
  - e.g. retrieve for each female employee a list of the names of her dependents

FEMALE_EMPS ← $\sigma_{SEX='F'}$(EMPLOYEE)

EMPNAMES ← $\Pi_{FNAME,LNAME,SSN}$(FEMAL_EMPS)

EMP_DEPENTS ←EMPNAMES x DEPENDENT

ACTUAL_DEPENDENT ← $\sigma_{SSN=ESSN}$(EMP_DEPENDENT)

RESULT← $\Pi_{FNAME,LNAME,DEPENDENT\_NAME}$(ACTUAL_DEPENDENT)

SQL

WHERE

SELECT

JOIN

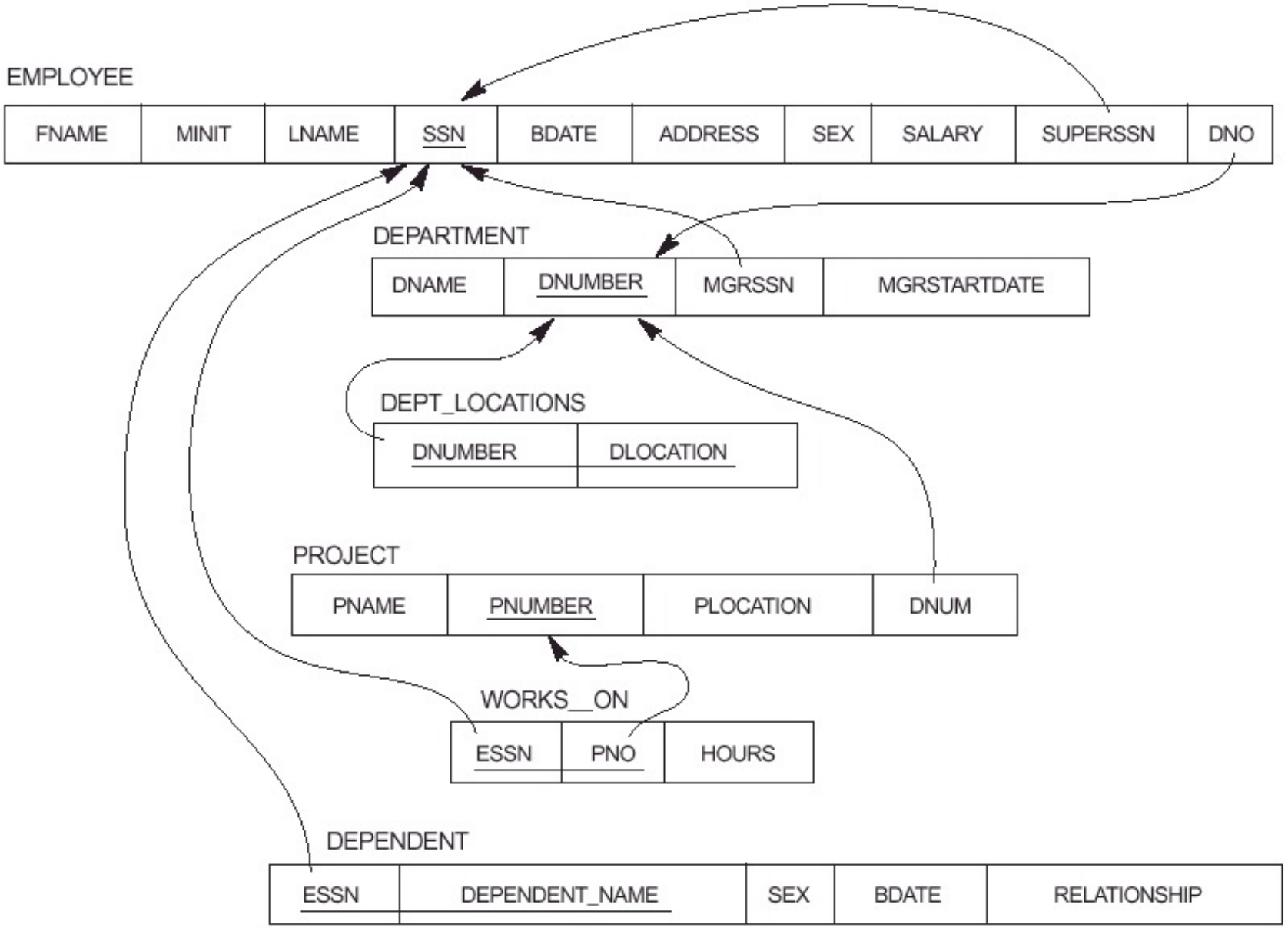
SELECT

EMPNAMES	FNAME	LNAME	SSN
	Alicia	Zelaya	999887777
	Jennifer	Wallace	987654321
	Joyce	English	453453453

EMP_DEPENTS ←EMPNAMES X DEPENDENT

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

EMP_DEPENDENTS	FNAME	LNAME	SSN	ESSN	DEPENDENT_NAME	SEX	BDATE	• • •
	Alicia	Zelaya	999887777	333445555	Alice	F	1986-04-05	• • •
	Alicia	Zelaya	999887777	333445555	Theodore	M	1983-10-25	• • •
	Alicia	Zelaya	999887777	333445555	Joy	F	1958-05-03	• • •
	Alicia	Zelaya	999887777	987654321	Abner	M	1942-02-28	• • •
	Alicia	Zelaya	999887777	123456789	Michael	M	1988-01-04	• • •
	Alicia	Zelaya	999887777	123456789	Alice	F	1988-12-30	• • •
	Alicia	Zelaya	999887777	123456789	Elizabeth	F	1967-05-05	• • •
	Jennifer	Wallace	987654321	333445555	Alice	F	1986-04-05	• • •
	Jennifer	Wallace	987654321	333445555	Theodore	M	1983-10-25	• • •
	Jennifer	Wallace	987654321	333445555	Joy	F	1958-05-03	• • •
	Jennifer	Wallace	987654321	987654321	Abner	M	1942-02-28	• • •
	Jennifer	Wallace	987654321	123456789	Michael	M	1988-01-04	• • •
	Jennifer	Wallace	987654321	123456789	Alice	F	1988-12-30	• • •
	Jennifer	Wallace	987654321	123456789	Elizabeth	F	1967-05-05	• • •
	Joyce	English	453453453	333445555	Alice	F	1986-04-05	• • •
	Joyce	English	453453453	333445555	Theodore	M	1983-10-25	• • •
	Joyce	English	453453453	333445555	Joy	F	1958-05-03	• • •
	Joyce	English	453453453	987654321	Abner	M	1942-02-28	• • •
	Joyce	English	453453453	123456789	Michael	M	1988-01-04	• • •
	Joyce	English	453453453	123456789	Alice	F	1988-12-30	• • •
	Joyce	English	453453453	123456789	Elizabeth	F	1967-05-05	• • •


EMPLOYEE
FNAME
MINIT
LNAME
SSN
BDATE
ADDRESS
SEX
SALARY
SUPERSSN
DNO
DEPARTMENT
DNAME
DNUMBER
MGRSSN
MGRSTARTDATE
DEPT_LOCATIONS
DNUMBER
DLOCATION
PROJECT
PNAME
PNUMBER
PLOCATION
DNUM
WORKS_ON
ESSN
PNO
HOURS
DEPENDENT
ESSN
DEPENDENT_NAME
SEX
BDATE
RELATIONSHIP

EMPLOYEE	FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
	John		Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
	Franklin		Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
	Alicia		Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
	Jennifer		Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
	Ramesh		Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
	Joyce		English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
	Ahmad		Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
	James		Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	null	1

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

FEMALE_EMPS ← $\sigma_{SEX='F'}$(EMPLOYEE)

EMPLOYEE	FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
	John		Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
	Franklin		Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
	Alicia		Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
	Jennifer		Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
	Ramesh		Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
	Joyce		English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
	Ahmad		Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
	James		Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	null	1

FEMALE_EMPS	FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
	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
	Joyce	A	English	453453453	1972-07-31	5631 Rice,Houston,TX	F	25000	333445555	5

EMPNAMES ← Π FNAME,LNAME,SSN(FEMAL_EMPS)

FEMALE_EMPS	FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
	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
	Joyce	A	English	453453453	1972-07-31	5631 Rice,Houston,TX	F	25000	333445555	5

EMPNAMES	FNAME	LNAME	SSN
	Alicia	Zelaya	999887777
	Jennifer	Wallace	987654321
	Joyce	English	453453453

EMPNAMES	FNAME	LNAME	SSN
	Alicia	Zelaya	999887777
	Jennifer	Wallace	987654321
	Joyce	English	453453453

EMP_DEPENTS ←EMPNAMES X DEPENDENT

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

EMP_DEPENDENTS	FNAME	LNAME	SSN	ESSN	DEPENDENT_NAME	SEX	BDATE	• • •
	Alicia	Zelaya	999887777	333445555	Alice	F	1986-04-05	• • •
	Alicia	Zelaya	999887777	333445555	Theodore	M	1983-10-25	• • •
	Alicia	Zelaya	999887777	333445555	Joy	F	1958-05-03	• • •
	Alicia	Zelaya	999887777	987654321	Abner	M	1942-02-28	• • •
	Alicia	Zelaya	999887777	123456789	Michael	M	1988-01-04	• • •
	Alicia	Zelaya	999887777	123456789	Alice	F	1988-12-30	• • •
	Alicia	Zelaya	999887777	123456789	Elizabeth	F	1967-05-05	• • •
	Jennifer	Wallace	987654321	333445555	Alice	F	1986-04-05	• • •
	Jennifer	Wallace	987654321	333445555	Theodore	M	1983-10-25	• • •
	Jennifer	Wallace	987654321	333445555	Joy	F	1958-05-03	• • •
	Jennifer	Wallace	987654321	987654321	Abner	M	1942-02-28	• • •
	Jennifer	Wallace	987654321	123456789	Michael	M	1988-01-04	• • •
	Jennifer	Wallace	987654321	123456789	Alice	F	1988-12-30	• • •
	Jennifer	Wallace	987654321	123456789	Elizabeth	F	1967-05-05	• • •
	Joyce	English	453453453	333445555	Alice	F	1986-04-05	• • •
	Joyce	English	453453453	333445555	Theodore	M	1983-10-25	• • •
	Joyce	English	453453453	333445555	Joy	F	1958-05-03	• • •
	Joyce	English	453453453	987654321	Abner	M	1942-02-28	• • •
	Joyce	English	453453453	123456789	Michael	M	1988-01-04	• • •
	Joyce	English	453453453	123456789	Alice	F	1988-12-30	• • •
	Joyce	English	453453453	123456789	Elizabeth	F	1967-05-05	• • •

$$ACTUAL\_DEPENDENT \leftarrow \sigma_{SSN=ESSN}(EMP\_DEPENDENT)$$

$$RESULT \leftarrow P_{FNAME,LNAME,DEPENDENT\_NAME}(ACTUAL\_DEPENDENT)$$

EMP_DEPENDENTS	FNAME	LNAME	SSN	ESSN	DEPENDENT_NAME	SEX	BDATE	• • •
	Alicia	Zelaya	999887777	333445555	Alice	F	1986-04-05	• • •
	Alicia	Zelaya	999887777	333445555	Theodore	M	1983-10-25	• • •
	Alicia	Zelaya	999887777	333445555	Joy	F	1958-05-03	• • •
	Alicia	Zelaya	999887777	987654321	Abner	M	1942-02-28	• • •
	Alicia	Zelaya	999887777	123456789	Michael	M	1988-01-04	• • •
	Alicia	Zelaya	999887777	123456789	Alice	F	1988-12-30	• • •
	Alicia	Zelaya	999887777	123456789	Elizabeth	F	1967-05-05	• • •
	Jennifer	Wallace	987654321	333445555	Alice	F	1986-04-05	• • •
	Jennifer	Wallace	987654321	333445555	Theodore	M	1983-10-25	• • •
	Jennifer	Wallace	987654321	333445555	Joy	F	1958-05-03	• • •
	Jennifer	Wallace	987654321	987654321	Abner	M	1942-02-28	• • •
	Jennifer	Wallace	987654321	123456789	Michael	M	1988-01-04	• • •
	Jennifer	Wallace	987654321	123456789	Alice	F	1988-12-30	• • •
	Jennifer	Wallace	987654321	123456789	Elizabeth	F	1967-05-05	• • •
	Joyce	English	453453453	333445555	Alice	F	1986-04-05	• • •
	Joyce	English	453453453	333445555	Theodore	M	1983-10-25	• • •
	Joyce	English	453453453	333445555	Joy	F	1958-05-03	• • •
	Joyce	English	453453453	987654321	Abner	M	1942-02-28	• • •
	Joyce	English	453453453	123456789	Michael	M	1988-01-04	• • •
	Joyce	English	453453453	123456789	Alice	F	1988-12-30	• • •
	Joyce	English	453453453	123456789	Elizabeth	F	1967-05-05	• • •

ACTUAL_DEPENDENTS	FNAME	LNAME	SSN	ESSN	DEPENDENT_NAME	SEX	BDATE
	Jennifer	Wallace	987654321	987654321	Abner	M	1942-02-28

RESULT	FNAME	LNAME	DEPENDENT_NAME
	Jennifer	Wallace	Abner

Join Operation

- Join
  - Similar to the Cartesian Product followed by a Select
  - Join condition.
  - e.g. retrieve the department name and the department manager name

$$DEPT\_MGR \leftarrow DEPARTMENT \infty_{MGRSSN=SSN} EMPLOYEE$$

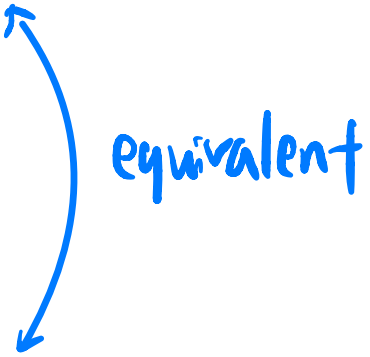
$$RESULT \leftarrow \Pi_{DNAME,LNAME,FNAME}(DEPT\_MGR)$$

- Without Join operation

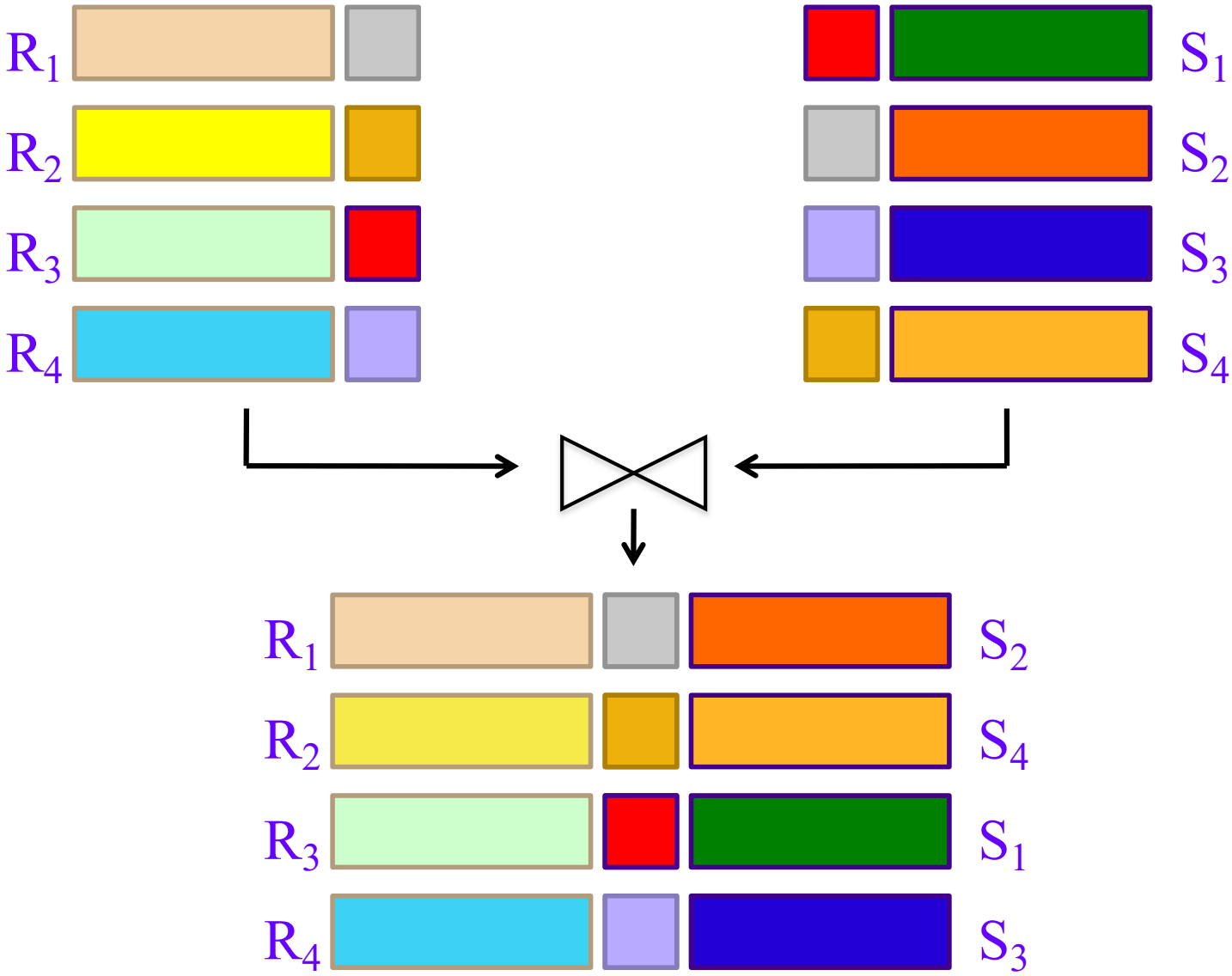
$$EMP\_DEPENDENTS \leftarrow DEPARTMENT \times EMPLOYEE$$

$$ACTUAL\_DEPENDENTS \leftarrow \sigma_{SSN=ESSN}(EMP\_DEPENDENT)$$

$$RESULT \leftarrow \Pi_{DNAME,LNAME,FNAME}(ACTUAL\_DEPENDENTS)$$



Relational Join



An SQL query walks into a bar and sees two tables.
He walks up to them and says "Can I join you?"

DEPT_MGR ← DEPARTMENT ∞ $_{MGRSSN=SSN}$ EMPLOYEE

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

EMPLOYEE	FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
	John		Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
	Franklin		Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
	Alicia		Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
	Jennifer		Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
	Ramesh		Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
	Joyce		English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
	Ahmad		Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
	James		Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	null	1

DNAME	DNUMBER	MGRSSN	• • •	FNAME	MINIT	LNAME	SSN	• • •
Research	5	333445555	• • •	Franklin	T	Wong	333445555	• • •
Administration	4	987654321	• • •	Jennifer	S	Wallace	987654321	• • •
Headquarters	1	888665555	• • •	James	E	Borg	888665555	• • •

Relational Algebra (cont.)

- Relational algebra
  - Two groups of operations
    - Set operations from mathematical set theory
      - Union
      - Intersection
      - Set Difference
      - Cartesian Product
    - Specially for RDB Operations
      - Select
      - Project
      - Join

Complete Set of Relational Algebra Operation

- Complete set of relational algebra operations
  - All the operations can be described by a sequence of operations of the set {σ, π, ∪, - , X }
  - e.g. ∩ 可以表示成： $R \cap S \equiv (R \cup S) - ((R-S) \cup (S-R))$
  - e.g. ∞_c (JOIN) 可以表示成： $R \infty_c S \equiv \sigma_c (R \times S)$
- Relational complete:
  - Any query language equivalent to the complete set

Additional Relational Operations

- AGGREGATE FUNCTIONS
  - Functions such as *SUM*, *COUNT*, *AVERAGE*, *MIN*, *MAX* are often applied to sets of values or sets of tuples in database applications
  - <grouping attributes> *F* <function list> (R)
  - The grouping attributes are optional
  - e.g. Retrieve the average salary of all employees (no grouping needed):

    R(AVGSAL) ← $F_{AVERAGE\ SALARY}$ (EMPLOYEE)
  - e.g. For each department, retrieve the department number, the number of employees, and the average salary (in the department):

F

R

grouping attributes

$$DNO\ \mathfrak{F}_{COUNT\ SSN,AVERAGE\ SALARY}(EMPLOYEE)$$

function list

EMPLOYEE	FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
	John		Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
	Franklin		Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
	Alicia		Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
	Jennifer		Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
	Ramesh		Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
	Joyce		English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
	Ahmad		Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
	James		Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	null	1

COUNT_SSN	AVERAGE_SALARY
8	35125

$$R(\text{DNO, NO\_OF\_EMPLOYEES, AVERAGE\_SAL}) \leftarrow {}_{\text{DNO}}\Im_{\text{COUNT SSN, AVERAGE SALARY}}(\text{EMPLOYEE})$$

rename attributes

EMPLOYEE	FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
	John		Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
	Franklin		Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
	Alicia		Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
	Jennifer		Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
	Ramesh		Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
	Joyce		English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
	Ahmad		Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
	James		Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	null	1

	DNO	NO_OF_EMPLOYEES	AVERAGE_SAL
	5	4	33250
	4	3	31000
	1	1	55000

Recursive Closure Operations

- Finding recursive relationship using looping mechanism

  e.g. find the employee supervised by Borg directly or indirectly

  1. Find the employee supervised by Borg directly (level 1)：

     $BORG\_SSN \leftarrow \Pi_{SSN}(\sigma_{FNAME='James' \text{ AND } LNAME='Borg'}(EMPLOYEE))$

     $SUPERVISION(SSN1, SSN2) \leftarrow \Pi_{SSN,SUPERSSN}(EMPLOYEE)$

     $RESULT1(SSN) \leftarrow \Pi_{SSN1}(SUPERVISION \; X_{SSN2=SSN} \; BORG\_SSN)$

  2. Find the employee supervised by Borg indirectly by level 2：

     $RESULT2(SSN) \leftarrow \Pi_{SSN1}(SUPERVISION \; X_{SSN2=SSN} \; RESULT1)$

  3. UNION result 1 & 2

     $RESULT \leftarrow RESULT2 \cup RESULT1$

- Transitive Closure (dynamic programming)

EMPLOYEE	FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO
	John		Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
	Franklin		Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
	Alicia		Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
	Jennifer		Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
	Ramesh		Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
	Joyce		English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
	Ahmad		Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
	James		Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	null	1

SUPERVISION	SSN1	SSN2
	123456789	333445555
	333445555	888665555
	999887777	987654321
	987654321	888665555
	666884444	333445555
	453453453	333445555
	987987987	987654321
	888665555	null

RESULT1(SSN) ←

Π_{SSN1}(SUPERVISION $X_{SSN2=SSN}$ BORG_SSN)

RESULT2(SSN) ←

Π_{SSN1}(SUPERVISION $X_{SSN2=SSN}$ RESULT1)

RESULT 1	SSN
	333445555
	987654321

(Supervised by Borg)

RESULT 2	SSN
	123456789
	999887777
	666884444
	453453453
	987987987

(Supervised by Borg's subordinates)

RESULT	SSN
	123456789
	999887777
	666884444
	453453453
	987987987
	333445555
	987654321

(RESULT1 ∪ RESULT2)

Relational Calculus

Relational Calculus

- Non-procedure language
  - Declarative expression to specify a retrieval request
  - No description of how to evaluate a query
  - Specifies what is to be retrieved rather than how to retrieve it
  - A calculus expression may be written in different ways, but the way it is written has no bearing on how a query should be evaluated.
- Expressive power of relational calculus & algebra is identical
- Relational Complete
  - a relational query language L is relationally complete if we can express in L any query that can be expressed in relational calculus.

Tuple Variables & Range Relations

- {t | COND(t) }
  - set of all tuples t that satisfy COND(t)
  - t: tuple variable (a row in a table)
  - COND(t): conditional expression involving t
  - e.g.

{t.Fname, t.Lname | Employee(t) and t.salary > 50000}

P(t) : predicate

Tuple Variables & Range Relations (cont.)

- $\{ t_1.A_j, t_2.A_k, \ldots, t_n.A_m \mid COND(t_1, t_2, \ldots, t_n, t_{n+1}, t_{n+2}, \ldots, t_{n+m}) \}$
  - t_i : tuple variable
  - A_i : attribute of relation on which ranges
  - COND: well-form formula (condition) of tuple relational calculus
  - Formula: made up of predicate calculus atoms connected via logical operators AND, OR, NOT
    - Every atom is a formula
    - If F1 & F2 are formulas then so are

      (F1 AND F2), (F1 OR F2), NOT(F1), NOT (F2)
  - Atoms: can be one of the following
    - $R(t_i)$, R: relation name, t_i: tuple variable
    - $t_i.A$ OP $t_j.B$: OP: comparison operators,

      t_i, t_j.:tuple variable, A, B: attributes
    - $t_i.A$ OP c: c: constant

↘ propositional calculus + quantifiers +

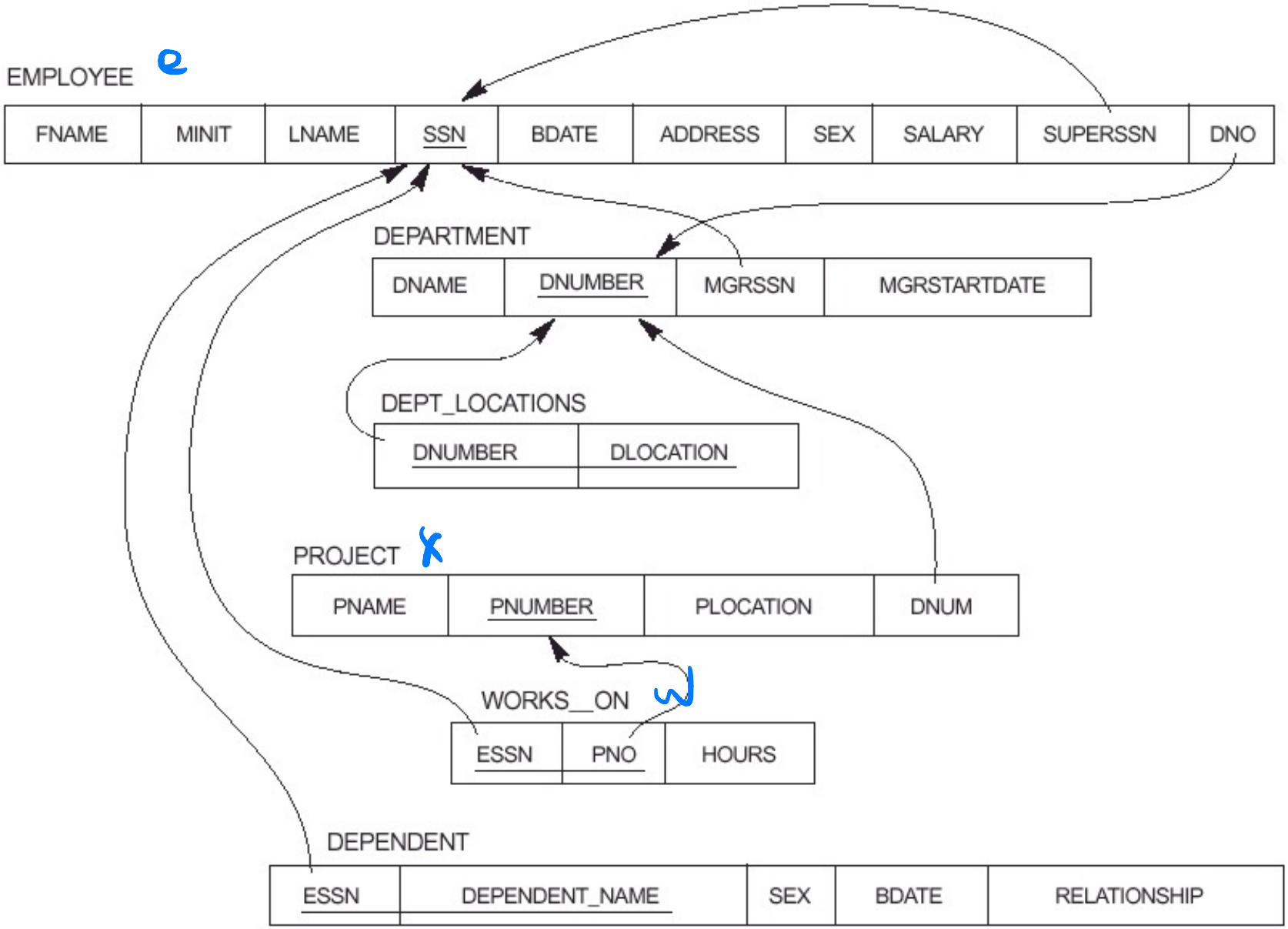
Existence & Universal Quantifiers

- Existence quantifier: ∃
- Universal quantifier: ∀
- Free & bound tuple variable
  - A tuple variable t is bound if it is quantified (appears in an (∃ t) or (∀ t) clause

e.g. tuple variable d is free in F1 & F2, bound in F3,

t is bound in F2

- F1: d.dname='research'
- F2:(∃ t)(d.dnumber = t.Dno)
- F3:(∀ d)(d.MgrSSN='333445555')


EMPLOYEE
FNAME
MINIT
LNAME
SSN
BDATE
ADDRESS
SEX
SALARY
SUPERSSN
DNO
DEPARTMENT
DNAME
DNUMBER
MGRSSN
MGRSTARTDATE
DEPT_LOCATIONS
DNUMBER
DLOCATION
PROJECT
PNAME
PNUMBER
PLOCATION
DNUM
WORKS_ON
ESSN
PNO
HOURS
DEPENDENT
ESSN
DEPENDENT_NAME
SEX
BDATE
RELATIONSHIP

Existence & Universal Quantifiers (cont.)

- Query 1: Retrieve the nsame & address of all employees who work for the 'Research' department

{ t.FName, t.Lname, t.Address | Employee(t) AND

(∃d) (DEPARTMENT(d) and d.DName='Research' and

d.Dnumber=t.DNo)

}

- Query 2: For every project located in 'Stafford', list the project number, the controlling department number, and the department manager's last name, birth date, and address

{ p.Pnumber, p.Dnum, m.Lname, m.Bdate, m.Address | Project(p) AND

Employee(m) AND p.Plocation='Stafford' AND

(∃ (d) (Department(d) AND p.Dnum=d.DNumber AND

d.MgrSSN=m.SSN)) }

Existence & Universal Quantifiers (cont.)

- Query 3': Find the name of each employee who works on some project controlled by department number 5

  { e.LName, e.FName| Employee(e) AND ((∃x)(∃w) (Project(x) AND Works_On(w) AND x.Dnum=5 AND w.ESSN=e.SSN AND x.Pnumber=w.Pno }

- Query 4: Make a list of project numbers for projects that involve an employee whose last name is 'Smith', either as a worker or as manager of the controlling department for the project

  { p.PNumber | Project(p) AND (

  worker

  ((∃e) (∃w) (Employee(e) AND Works_on(w) AND w.Pno=p.PNumber AND e.LName='Smith' AND e.SSN=w.ESSN)) or

  ((∃m) (∃d) (employee(m) AND Department(d) AND p.Dnum=d.DNUmber AND d.MgrSSN=m.SSN AND m.Lname='Smith'))

) }

  manager

Existence & Universal Quantifiers (cont.)

- Query 3': List the name of each employee who works on some project controlled by department number 5

  { e.LName, e.FName| Employee(e) AND ((∃x)(∃w) (Project(x) AND Works_On(w) AND x.Dnum=5 AND w.ESSN=e.SSN AND x.Pnumber=w.Pno }

- Query 3: List the name of each employee who works on all projects controlled by department number 5

  { e.LName, e.FName| Employee(e) AND

  ((∀ x) (NOT(Project(x) OR NOT (x.Dnum=5) OR ((∃w) (Works_On(w) AND w.ESSN=e.SSN AND x.Pnumber=w.Pno)))) }

Domain Relational Calculus

- SQL
  - Based on tuple relational calculus
  - developed by IBM Research at San Jose
- QBE (Query-By-Example)
  - Related to domain relational calculus
  - developed by IBM Research at Yorktown Heights
  - First graphical query language with minimum syntax developed for database systems
  - Microsoft Access (Office)
- Any query language that can be expressed in the relational algebra can also be expressed in the domain or tuple relational calculus

GUI

Domain Relational Calculus (cont.)

- Variables range over single values from domain of attributes, rather than having variables range over tuples
- $\{x_1, x_2, \ldots, x_n | \text{COND}(x_1, x_2, \ldots, x_n, x_{n+1}, x_{n+2}, \ldots, x_{n+m})\}$
  - x_i: domain variables that range over domain (of attributes)
  - COND: well-form formula (condition) of domain relational calculus
  - Formula: made up of atoms connected via logical operators AND, OR, NOT
  - Every atom is a formula
  - If F1 & F2 are formulas then so are (f1 And F2), (F1 OR F2) NOT(F1), NOT (F2)
  - Atoms: can be one of the following
    - R(xi), R: relation name, xi: domain variable
    - x_i OP x_j: OP: comparison operators, x_i, x_j.:domain variable
    - x_i OP c: c:constant

Examples of Domain Relational Calculus

- Query0: Retrieve the birth date & address of the employee whose name is 'John B. Smith'
  - {u,v | (∃ q)(∃ r)(∃ s)(∃ t)(∃ w)(∃ x)(∃ y)(∃ z) (Employee(q,r,s,t,u,v,w,x,y,z) AND q='John' AND r='B' AND s='Smith') }
- Query 1: Retrieve the name & address of all employees who work for the 'Research' department
  - { q,s,v | (∃ z)(∃ l)(∃ m) Employee(q,r,s,t,u,v,w,x,y,z) AND Department(l,m,n,o) AND l='Research' AND m=z) }
- Query 2: For every project located in 'Stafford', list the project number, the controlling department number, and the department manager's last name, birth date, and address
  - {i,k,s,u,v | (∃ j)(∃ m)(∃ n)(∃ t) Project(h,i,j,k) AND Employee(q,r,s,t,u,v,w,x,y,z) AND Department(l,m,n,o) AND k=m AND n=t AND j='Stafford')}

Summary

- Relational algebra (How)
  - Set operations from mathematical set theory
    - Union
    - Intersection
    - Set Difference
    - Cartesian Product
  - Specially for RDB Operations
    - Select
    - Project
    - Join
- Relational Calculus (What)
    - Tuple relational calculus (SQL)
    - Domain relational calculus (QBE)