Chapter 4
Relational Algebra

Content

- Introduction
- Relational algebra
- Set operations on relations
- Selection
- Projection
- Cartesian production
- Join operation
- Division operation
- Other operations
- Update operations

Introduction

- Consider manipulations on the relation EMPLOYEE
  - Add a new employee
  - Move the employee whose name “Tung” to department 1
  - List names and birth dates of employees whose salary are over 20000

FName	LName	BirthDate	Address	Sex	Salary	DNo
Tung	Nguyen	12/08/1955	638 NVC Q5	Nam	40000	5 1
Hang	Bui	07/19/1968	332 NTH Q1	Nu	25000	4
Nhu	Le	06/20/1951	291 HVH QPN	Nu	43000	4
Hung	Nguyen	09/15/1962	Ba Ria VT	Nam	38000	5
Quang	Pham	11/10/1937	450 TV HN	Nam	55000	1

Introduction

- Study database programming
  - How the user can ask queries of the database
    - Select
  - How the user can modify the contents of the database
    - Insert, delete and update

- Relational model
  - Relational Algebra
    - Present a query by expressions
  - Relational Calculus
    - Present the result of a query
  - SQL (Structured Query Language)

Review

- Algebra
  - Operators
  - Atomic operands

- In algebra arithmetic
  - Operators : +, -, *, /
  - Operand – Variable : x, y, z
  - Constant
  - Expression
    - (x+7) / (y-3)
    - (x+y)*z and/or (x+7) / (y-3)

Relational algebra

- Variables – Relations
  - Set
- Operators
  - Set operations
    - Union $\cup$
    - Intersection $\cap$
    - Difference $-$
  - Retrieve parts of a relation
    - Selection σ
    - Projection π
  - Combine tuples of two relations
    - Cartesian product $\times$
    - Join $\bowtie$

Relational algebra

- Constant
  - Instance of the relation
- Expression
  - A query
  - A sequence of relational algebra operations
- Operands and results of expressions
  - Sets

Content

- Introduction
- Relational algebra
- **Set operations**
- Selection
- Projection
- Cartesian product
- Join operation
- Division operation
- Other operations
- Update operations

Set operation

- Relation is a set of tuples
  - The union $R \cup S$
  - The intersection $R \cap S$
  - The difference $R - S$

- Union Compatibility
  - Two relation schemas $R(A_1, A_2, \ldots, A_n)$ and $S(B_1, B_2, \ldots, B_n)$ are union compatibility if
    - The same degree n
    - And $DOM(A_i)=DOM(B_i)$, $1\leq i \leq n$

- The result of $\cup$, $\cap$, and $-$ operations
  - Relation

Union

- Given two relations R & S that are union compatible
- The union of R and S
  - Notation $R \cup S$
  - A relation consists of tuples that are in R or S or both (an element appears only one)

$$R \cup S = \{ t \mid t \in R \lor t \in S \}$$

- Example

R	A	B
	α	1
	α	2
	β	1

S	A	B
	α	2
	β	3

R ∪ S	A	B
	α	1
	α	2
	β	1
	~~α~~	~~2~~
	β	3

Intersection

- Given two relations R & S that are union compatible
- The intersection of R and S
  - Denotation $R \cap S$
  - A relation consists of tuples that are in R and S

$$R \cap S = \{ t \,/\, t \in R \wedge t \in S \}$$

- Example

R	A	B
	α	1
	α	2
	β	1

S	A	B
	α	2
	β	3

R ∩ S	A	B
	α	2

Difference

- Given two relations R & S that are union compatible
- The difference of R and S
  - Denotation R − S
  - A relation consists of tuples that are in R but not in S

$$R - S = \{ t \ / \ t \in R \land t \notin S \}$$

- Example

R	A	B
	α	1
	α	2
	β	1

S	A	B
	α	2
	β	3

R − S	A	B
	α	1
	β	1

Properties

- Commutative

$$R \cup S = S \cup R$$

$$R \cap S = S \cap R$$

- Associative

$$R \cup (S \cup T) = (R \cup S) \cup T$$

$$R \cap (S \cap T) = (R \cap S) \cap T$$

Content

- Introduction
- Relational algebra
- Set operations
- **Selection**
- Projection
- Cartesian product
- Join operation
- Division operation
- Other operations
- Queries in relational algebra

Selection

- Is applied to relation R to produce a new relation with a subset of R's tuples
- Tuples in the resulting relation satisfy some condition C
- Denotation $\sigma_{C}(R)$

- C is a Boolean expression made up of <u>clauses</u>
  - <attribute> <comparison operator> <constant>
  - <attribute> <comparison operator> <attribute>

    - <comparison op> : $<$, $>$, $\leq$, $\geq$, $\neq$, $=$
    - Clauses are connected by Boolean operator : $\wedge$, $\vee$, $\neg$

Selection

- The result is a relation
  - The same list of attributes as R
  - The number of tuples is less than or equal to the number of tuples of R

- Example

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

Selection

- Selection operator is commutative

$$\sigma_{C1}(\sigma_{C2}(R)) = \sigma_{C2}(\sigma_{C1}(R)) = \sigma_{C1 \wedge C2}(R)$$

Example 1

- List all employees who work in department 4
  - Relation: EMPLOYEE
  - Attribute: DNo
  - Condition: DNo=4

$$\sigma_{DNo=4}(EMPLOYEE)$$

Example 2

- Select tuples for all employees who either work in department 4 and make over $25,000 per year or work in department 5 and make over $30,000
  - Relation: EMPLOYEE
  - Attributes: SALARY, DNO
  - Condition:
    - (SALARY>25000 and DNO=4) or
    - (SALARY>30000 and DNO=5)

$$\sigma_{(SALARY>25000 \wedge DNO=4) \vee (SALARY>30000 \wedge DNO=5)}(EMPLOYEE)$$

Content

- Introduction
- Relational algebra
- Set operations
- Selection
- **Projection**
- Cartesian product
- Join operation
- Division operation
- Other operations
- Update operations

Projection

- Is used to produce from a relation R a new relation that has only some of R's columns
- Denotation $\pi_{A1, A2, \ldots, Ak}(R)$
- The result is a relation
  - Has k attributes
  - The number of tuples is less than or equal to the number of tuples of R
- Example

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

Projection

- Projection operator is not commutative

$$\pi_{X,Y}(R) = \pi_X(\pi_Y(R))$$

$$\pi_{A1,\ A2,\ \ldots,\ An}(\pi_{A1,\ A2,\ \ldots,\ Am}(R)) = \pi_{A1,\ A2,\ \ldots,\ An}(R)\ ,\ n \leq m$$

Example 3

- List out the name and salary of employees

$$\pi_{\text{LNAME, FNAME, SALARY}} (\text{EMPLOYEE})$$

Example 4

- Find the SSN of employees who either work on projects or have dependents

Example 5

- Find the SSN of employees who work on projects and have dependents

Example 6

- Find the SSN of employees who do not have any dependents

Extended projection

- Extending the projection operator to allow it to compute with components of tuples
- Denotation $\pi_{F1,\ F2,\ \ldots,\ Fn}$ (E)
  - E is a relation algebra expression
  - $F_1, F_2, \ldots, F_n$ are arithmetic expressions involving
    - Attributes in E
    - Constants
    - Arithmetic operators (a + b : sum)
    - String operators (c || d : concatenate)

Example 7

- List out the employees' name and salary increased by 10%

$$\pi_{\text{LNAME, FNAME, SALARY*1.1}} \text{(EMPLOYEE)}$$

Sequences of operations

- Apply several relational algebra operations one after one
  - A single relational algebra expression

$$\pi_{A1,\ A2,\ \ldots,\ Ak}(\sigma_C(R)) \qquad \sigma_C(\pi_{A1,\ A2,\ \ldots,\ Ak}(R))$$

  - Break down a complex expression into simpler steps
    - Step 1 $\sigma_C(R)$
    - Step 2 $\pi_{A1,\ A2,\ \ldots,\ Ak}$(the result of step 1)

Giving a name

Assignment operator

- Is often used to receive the result of an operation
  - The intermediate result in a sequence of operations
- Denotation $\leftarrow$

- Example
  - Step 1: $S \leftarrow \sigma_{C}(R)$
  - Step 2:

$$RESULT \leftarrow \pi_{A1, A2, \ldots, Ak}(S)$$

Rename operator

- Is used to rename either the relation name or attribute name
  - Relation

    Examine R(B, C, D)

    $\rho_S(R)$: Rename the name of relation R to S

  - Attribute

    $\rho_{X, C, D}(R)$: Rename the name of attribute B to X

Rename the name of relation R to S and the name of attribute B to X

$$\rho_{S(X,C,D)}(R)$$

Example 8

- List out the name of employees who work in department 4

- Case 1: $\pi_{LNAME, FNAME}(\sigma_{DNO=4}(EMPLOYEE))$

- Case 2: $EMP\_DEP4 \leftarrow \sigma_{DNO=4}(EMPLOYEE)$

  $RESULT \leftarrow \pi_{LNAME, FNAME}(EMP\_DEP4)$

  $\rho_{RESULT(LASTNAME, FIRSTNAME)}(\pi_{LNAME, FNAME}(EMP\_DEP4))$

$RESULT(LASTNAME, FIRSTNAME) \leftarrow \pi_{LNAME, FNAME}(EMP\_DEP4)$

Content

- Introduction
- Relational algebra
- Set operations
- Selection
- Projection
- **Cartesian product**
- Join operation
- Division operation
- Other operations
- Update operations

Cartesian product

- Cross-product
  - Is used to combine tuples from two relations in a combinatorial fashion
- Denotation $R \times S$
- The result is a relation Q
  - Q has one tuple for each combination of tuples, one from R and one from S
  - If R has *u* tuples and S has *v* tuples,
    - Then Q will have $(u \times v)$ tuples
  - If R has *n* attributes and S has *m* attributes,
    - Then Q will have $(n + m)$ attributes $(R^+ \cap S^+ = \varnothing)$

Cartesian product

- Example

R	A	B
	α	1
	β	2

S	X	C	D
	α	10	+
	β	10	+
	β	20	-
	γ	10	-

$\rho_{(X,C,D)}(S)$

$R \times S$

A	B	X	C	D
α	1	α	10	+
α	1	β	10	+
α	1	β	20	-
α	1	γ	10	-
β	2	α	10	+
β	2	β	10	+
β	2	β	20	-
β	2	γ	10	-

Cartesian product

- Example

R	A	B
	α	1
	β	2

S	B	C	D
	α	10	+
	β	10	+
	β	20	-
	γ	10	-

unambiguous

$R \times S$	A	R.B	S.B	C	D
	α	1	α	10	+
	α	1	β	10	+
	α	1	β	20	-
	α	1	γ	10	-
	β	2	α	10	+
	β	2	β	10	+
	β	2	β	20	-
	β	2	γ	10	-

Cartesian product

- Cartesian product is often followed by a selection operation

$R \times S$

A	R.B	S.B	C	D
α	1	α	10	+
α	1	β	10	+
α	1	β	20	-
α	1	γ	10	-
β	2	α	10	+
β	2	β	10	+
β	2	β	20	-
β	2	γ	10	-

$\sigma_{A=S.B}(R \times S)$

A	R.B	S.B	C	D
α	1	α	10	+
β	2	β	10	+
β	2	β	20	-

Example 9

- For each department, list out the information of the manager

DNAME	DNUMBER	MGRSSN	MGRSTARTDATE
Nghien cuu	5	333445555	05/22/1988
Dieu hanh	4	987987987	01/01/1995
Quan ly	1	888665555	06/19/1981

DNAME	DNUMBER	MGRSSN	MGRSTARTDATE	SSN	FNAME	LNAME	…
Nghien cuu	5	333445555	05/22/1988	333445555	Tung	Nguyen	…
Dieu hanh	4	987987987	01/01/1995	987987987	Hung	Nguyen	…
Quan ly	1	888665555	06/19/1981	888665555	Vinh	Pham	…

SSN	FNAME	LNAME	BIRTHDATE	ADDRESS	SEX	SALARY	DNO
333445555	Tung	Nguyen	12/08/1955	638 NVC Q5	Nam	40000	5
999887777	Hang	Bui	07/19/1968	332 NTH Q1	Nu	25000	4
987654321	Nhu	Le	06/20/1951	291 HVH QPN	Nu	43000	4
987987987	Hung	Nguyen	09/15/1962	Ba Ria VT	Nam	38000	5

Example 9

- Step 1:
  - Cartesian product DEPARTMENT & EMPLOYEE

$$EMP\_DEP \leftarrow (DEPARTMENT \times EMPLOYEE)$$

- Step 2:
  - Select tuples that satisfy the condition MgrSSN=SSN

$$RESULT \leftarrow \sigma_{MGRSSN=SSN}(EMP\_DEP)$$

Example 10

- Find the highest salary in company

FNAME	LNAME	...	SALARY	...
Tung	Nguyen	...	40000	...
Hang	Bui	...	25000	...
Nhu	Le	...	43000	...
Hung	Nguyen	...	38000	...

...	SALARY	...
...	40000	...
...	25000	...
...	43000	...
...	38000	...

Example 10

- <u>Step 1:</u>
  - Select salaries which are not the highest one

$$R1 \leftarrow (\pi_{SALARY}(EMPLOYEE))$$

$$R2 \leftarrow \sigma_{EMPLOYEE.SALARY < R1.SALARY}(EMPLOYEE \times R1)$$

$$R3 \leftarrow \pi_{EMPLOYEE.SALARY}(R2)$$

- <u>Step 2:</u>
  - Let do the difference of the set of salary and salary in R3

$$RESULT \leftarrow \pi_{SALARY}(EMPLOYEE) - R3$$

Example 11

- Find the departments that have the same locations as the department 5

Which locations does the department 5 have?

DNUMBER	DLOCATION
1	TP HCM
4	HA NOI
5	VUNGTAU
5	NHATRANG
5	TP HCM

Which departments will have locations which are in that set

DNUMBER	DLOCATION
1	TP HCM
4	HA NOI
5	VUNGTAU
5	NHATRANG
5	TP HCM

Example 11

- Step 1:
  - Find the locations of the department 5

$$LOC\_DEP5(Loc) \leftarrow \pi_{DLOCATION}(\sigma_{DNUMBER=5}(DEPT\_LOCATIONS))$$

- Step 2:
  - Select the departments that have the same locations as LOC_DEP5

$$R1 \leftarrow \sigma_{DNUMBER \neq 5}(DEPT\_LOCATIONS)$$

$$R2 \leftarrow \sigma_{DLOCATION=LOC}(R1 \times LOC\_DEP5)$$

$$RESULT \leftarrow \pi_{DNUMBER}(R2)$$

Content

- Introduction
- Relational algebra
- Set operations
- Selection
- Projection
- Cartesian product
- **Join operation**
  - Natural join
  - Theta join
  - Equi join
- Divide operation
- Other operations
- Update operations

Join operation

- Is used to combine related tuples from 2 relations into single tuples
- Denotation R $\bowtie$ S
  - $R(A_1, A_2, \ldots, A_n)$ and $S(B_1, B_2, \ldots, B_m)$

- Result is a relation Q
  - Has *(n + m)* attributes $Q(A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_m)$
  - A tuple of Q is a combination of tuples from R and S satisfying some join condition
    - The form : $A_i \ \theta \ B_j$
    - A_i : the attribute from R, B_j : the attribute from S
    - A_i and B_j have the same domain
    - θ : comparison operators $\neq, =, <, >, \leq, \geq$

Join operation

- Categories
  - ***Theta join*** pairs tuples using one specific condition
    - Denotation $R \bowtie_C S$
    - C refers to an arbitrary condition for attributes
  - ***Equijoin*** when C involves equality comparisons only
  - ***Natural join***
    - Denote $R \bowtie S$ or $R * S$
    - $R^+ \cap S^+ \neq \emptyset$
    - Only one join attribute is kept

Join operation

- Example of theta join

R	A	B	C
	1	2	3
	4	5	6
	7	8	9

S	D	E
	3	1
	6	2

$R \bowtie_{B<D} S$

A	B	C	D	E
1	2	3	3	1
1	2	3	6	2
4	5	6	6	2

$$R \bowtie_{C} S \quad = \quad \sigma_{C}(R \times S)$$

Join operation

- Example of equijoin

R	A	B	C
	1	2	3
	4	5	6
	7	8	9

S	D	E
	3	1
	6	2

$R \bowtie_{C=D} S$

A	B	C	D	E
1	2	3	3	1
4	5	6	6	2

R	A	B	C
	1	2	3
	4	5	6
	7	8	9

S	S.C	D
	3	1
	6	2

$\rho_{(S.C,D)} S$

$R \bowtie_{C=S.C} S$

A	B	C	S.C	D
1	2	3	3	1
4	5	6	6	2

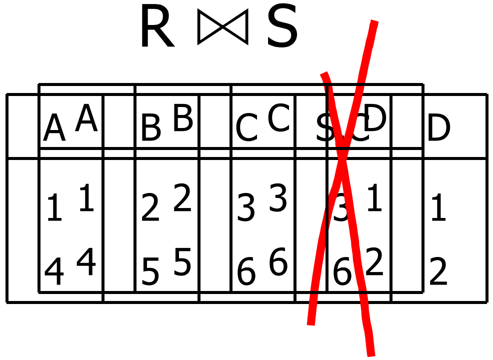
Join operation

- Example of natural join

R	A	B	C
	1	2	3
	4	5	6
	7	8	9

S	C	D
	3	1
	6	2

R ⋈ S



Example 12

- Find the employees whose salary are greater than the salary of the employee 'Tùng'

EMPLOYEE(LNAME, FNAME, SSN, ..., **SALARY**, DNO)

$R1(SAL) \leftarrow \pi_{SALARY}(\sigma_{FNAME='Tung'}(EMPLOYEE))$

$RESULT \leftarrow EMPLOYEE \bowtie_{SALARY>SAL} R1$

RESULT(LNAME, FNAME, SSN, ..., **SALARY**, DNO, **SAL**)

Example 13

- For each employee, find the information of the department that he/she is working for

EMPLOYEE(LNAME, FNAME, SSN, ..., **DNO**)

DEPARTMENT(DNAME, **DNUMBER**, MGRSSN, MGRSTARTDATE)

$$RESULT \leftarrow EMPLOYEE \bowtie_{DNO=DNUMBER} DEPARTMENT$$

RESULT(LNAME, FNAME, SSN, ..., **DNO**, DNAME, **DNUMBER**, ...)

Example 14

- Find the locations for each department

DEPARTMENT(DNAME, **DNUMBER**, MGRSSN, MGRSTARTDATE)

DEPT_LOCATIONS(**DNUMBER**, DLOCATION)

RESULT $\leftarrow$ DEPARTMENT $\bowtie$ DEPT_LOCATIONS

RESULT(DNAME, **DNUMBER**, MGRSSN, MGRSTARTDATE, DLOCATION)

Example 9

- For each department, list out the information of the manager

Example 10

- Find the highest salary in company

Example 11

- Find the departments that have the same locations as the department 5

A complete set of relational algebra operations

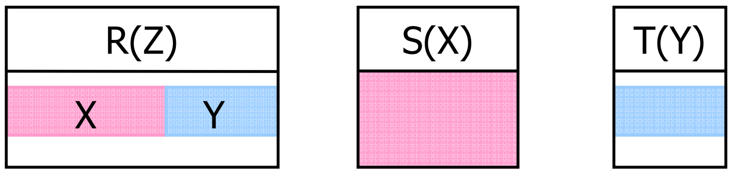
- The set of relational algebra operations $\{\sigma, \pi, \times, -, \cup\}$ is called a <u>complete set</u>
  - Any of other relational algebra operations can be expressed as a sequence of operations from this set
  - Example
    - $R \cap S = R - (R - S)$
    - $R \bowtie_C S = \sigma_C(R \times S)$

Content

- Introduction
- Relational algebra
- Set operations
- Selection
- Projection
- Cartesian product
- Join operation
- **Division operation**
- Other operations
- Update operations

Division

- Is used to retrieve tuples from R that satisfy <u>all</u> tuples from S
- Denotation R ÷ S
  - R(Z) and S(X)
    - Z is attribute set of R, X is attribute set of S
    - $X \subseteq Z$
- Result is a relation T(Y)
  - Has Y=Z-X
  - Includes tuples t, if for all $t_S \in S$, there exists a tuple $t_R \in R$ with 2 conditions
    - $t_R(Y) = t$
    - $t_R(X) = t_S(X)$



Division

- Example

R ÷ S

R	A	B	C	D	E
	α	a	α	a	1
	α	a	γ	a	1
	α	a	γ	b	1
	β	a	γ	a	1
	β	a	γ	b	3
	γ	a	γ	a	1
	γ	a	γ	b	1
	γ	a	β	b	1

S	D	E
	a	1
	b	1

A	B	C
α	a	γ
γ	a	γ

Example 15

- Find the SSN of employees who work on all the projects

Example 16

- Find the SSN of employees who work for all projects that the department 4 controls

Division

- Express the division operation by the complete set of relational algebra operations

$$Q1 \leftarrow \pi_Y (R)$$

$$Q2 \leftarrow Q1 \times S$$

$$Q3 \leftarrow \pi_Y(Q2 - R)$$

$$T \leftarrow Q1 - Q3$$

Content

- Introduction
- Relational algebra
- Set operations
- Selection
- Projection
- Cartesian product
- Join operation
- Division operation
- **Other operations**
  - Aggregation operators
  - Grouping
  - Outer join
- Update operations

Aggregation operators

- Input : the collections of values from the DB
- Output : a single value
- Include
  - AVG
  - MIN
  - MAX
  - SUM
  - COUNT

Aggregation operators

- Example

R	A	B
	1	2
	3	4
	1	2
	1	2

SUM(B) = 10

AVG(A) = 1.5

MIN(A) = 1

MAX(B) = 4

COUNT(A) = 4

Grouping

- Is used to consider a relation in groups, corresponding to the value of columns
- Denotation

$$_{G1,\ G2,\ \ldots,\ Gn}\mathcal{F}_{F1(A1),\ F2(A2),\ \ldots,\ Fn(An)}(E)$$

  - E is relational algebra expression
  - G1, G2, …, Gn : grouping attributes
  - F1, F2, …, Fn : aggregation operators
  - A1, A2, …, An : aggregated attributes

Groping

- Example

R	A	B	C
	α	2	7
	α	4	7
	β	2	3
	γ	2	10

$\mathcal{F}_{SUM(C)}(R)$

SUM_C
27

${}_{A}\mathcal{F}_{SUM(C)}(R)$

A	SUM_C
α	14
β	3
γ	10

Example 17

- The number of employees and the average salary of the company

Example 18

- For each department, find the number of employees and the average salary

Example 19

- Find the name of departments that have the largest number of employees

Example 20

- Find the name of employees who work the largest number of projects

Outer join

- Is used to avoid the loss of information
  - A theta join is taken first
  - Then, the tuples that failed to join with any tuple of the other relation are added to the result

- Three cases
  - Left outer join ⟕
  - Right outer join ⟖
  - Left and right outer join ⟗

Example 21

- List out the name of employees and the name of department that they are the manager if any

$$R1 \leftarrow EMPLOYEE ⟕_{SSN=MGRSSN} DEPARTMENT$$

$$RESULT \leftarrow \pi_{FNAME,LNAME,DNAME}(R1)$$

FNAME	LNAME	DNAME
Tung	Nguyen	Nghien cuu
Hang	Bui	null
Nhu	Le	null
Vinh	Pham	Quan ly

Example 22

- List out the name of departments and the number of employees of that department

If a department has just been established and not yet been arranged the employees, then what will be the result?

Content

- Introduciton
- Relational algebra
- Set operations
- Selection
- Projection
- Cartesian product
- Join operation
- Division operation
- Other operations
- **Update operations**

Update operations

- The content of the database can be updated by update operations
  - Insertion
  - Deletion
  - Update
- These operations are expressed by an assignment operation

$$R_{new} \leftarrow \text{operations on } R_{old}$$

Insertion operation

- Is expressed

$$R_{new} \leftarrow R_{old} \cup E$$

  - R is a relation
  - E is a relational algebra expression

- Example
  - Assign the employee whose SSN is 123456789 the project with SSN is 20 and the number of working hours is 10

$$WORKS\_ON \leftarrow WORKS\_ON \cup (\text{'123456789'}, 20, 10)$$

Deletion operation

- Is expressed

$$R_{new} \leftarrow R_{old} - E$$

  - R is a relation
  - E is a relational algebra expression

- Example
  - Delete all work assignments of the employee 123456789

$$WORKS\_ON \leftarrow WORKS\_ON - \sigma_{SSN='123456789'}(WORKS\_ON)$$

Example 23

- Remove work assignments that have locations in 'Ha Noi'

Update operation

- Is expressed

$$R_{new} \leftarrow \pi_{F1,\ F2,\ \ldots,\ Fn}\ (R_{old})$$

  - R is a relation
  - Fi is an arithmetic expression that results in the new value for attributes

- Example
  - Increase working hours to 1.5 times for all employees

$$WORKS\_ON \leftarrow \pi_{SSN,\ PNO,\ HOURS*1.5}(WORKS\_ON)$$

Example 24

- Increase working hours to 1.5 times for assignments that are over 30 hours, the remain will be increased to 2 times