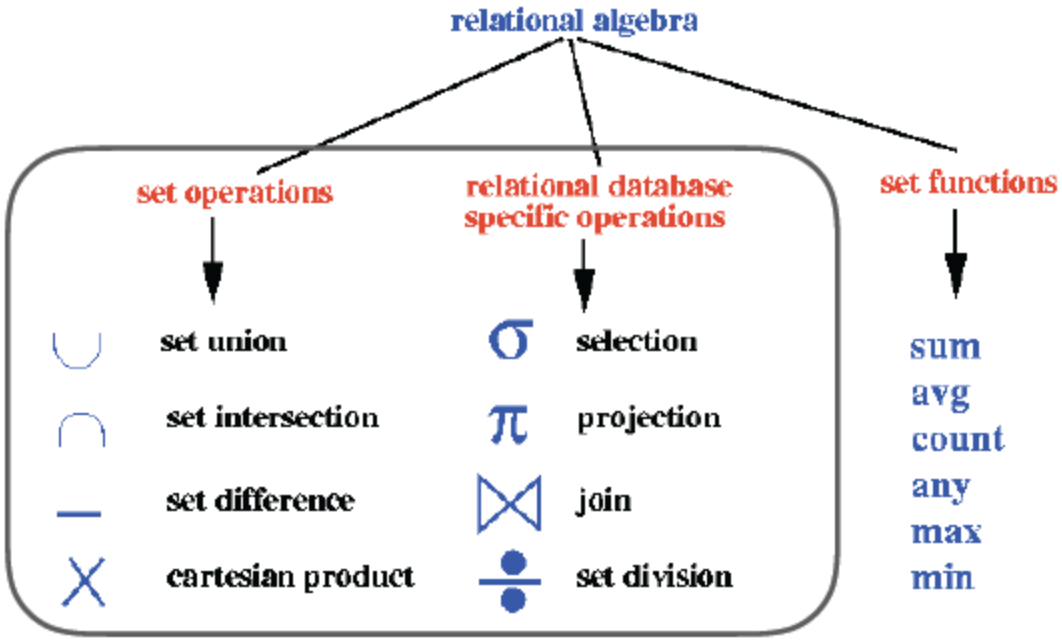
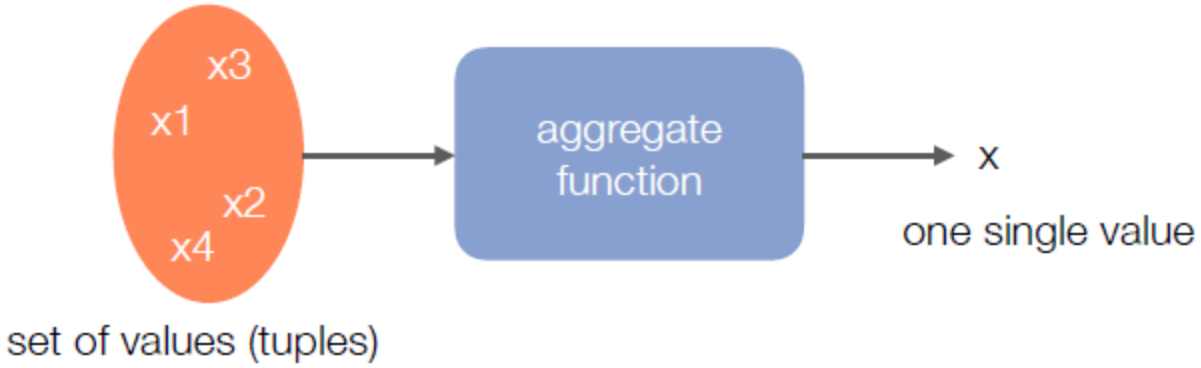
CS354: Database

Recap: Relational Algebra Part I



Set (Aggregate) Functions

- Operates on a set of values and produce a single value
- Can also be known as aggregate functions
- Common functions include SUM, AVERAGE, MAXIMUM, MINIMUM, and COUNT



Example: Set Functions

A = {1, 4, 5, 10, 15}

Function	Description	Value
sum(A)	sum of all values in the (numeric) set	35
avg(A)	average of all values in the (numeric) set	7
max(A)	maximum value of all values in the set	15
min(A)	minimum value of all values in the set	1
any(A)	TRUE if set is not empty, otherwise FALSE	TRUE
count(A)	cardinality (number of elements) of set	5

Additional Operations: Generalized Projection

- Allows functions of attributes to be included in the projection list

$$\pi_{f_1(a_1),f_2(a_2),\cdots,f_n(a_n)}(R)$$

- Examples:

$$\pi_{\text{LNAME,FNAME,SALARY}*1.03}(\text{EMPLOYEE})$$

$$\pi_{\text{SSN,FNAME,AGE}/2+7,\text{SEX}}(\text{EMPLOYEE})$$

Additional Operations: Group By Aggregate

- Groups are formed using one more attribute value(s)
- Aggregate functions applied independently to each group
- Examples:
  - How many people bought an iPad?
  - What is the average age of students in the Database Systems class?
  - What is the average salary of the different departments?

Example: Group By Aggregate

SSN	FName	Other	Sex	DNo	Salary
111-11-1111	John	…	M	4	50,000
242-12-2340	Mary	…	F	5	60,000
222-22-2222	James	…	M	5	80,000
333-33-3333	Jake	…	M	4	60,000

Group by DNO

111-11-1111	John	…	M	4	50,000
333-33-3333	Jake	…	M	4	60,000

avg(salary) = 55,0000

242-12-2340	Mary	…	F	5	60,000
222-22-2222	James	…	M	5	80,000

avg(salary) = 70,0000

Group By Aggregate Operation

- Notation: $a_1,a_2,\cdots,a_N \mathcal{F}_{f_1(a_1),f_2(a_2),\cdots,f_M(a_M)}(R)$
  - a_1, a_2, …, a_N = attributes used to form groups
  - $f_1(a_1)$, $f_2(a_2)$, …, $f_M(a_M)$ = set functions applied on each group
- Result is always a relation with the following attributes:
  - Grouping attributes (to differentiate the tuples)
  - Set function values (attributes named after function name)

a1	a2	…	aN	f1	f2	…	fM

Example: Group By Aggregate (2)

SSN	FName	Other	Sex	DNo	Salary
111-11-1111	John	…	M	4	50,000
242-12-2340	Mary	…	F	5	60,000
222-22-2222	James	…	M	5	80,000
333-33-3333	Jake	…	M	4	60,000

$$\text{DNO,Sex}\mathcal{F}_{\text{avg(Salary)},\text{count(SSN)}}(\text{EMPLOYEE})$$

DNo	Sex	Avg	Count
4	M	55,000	2
5	M	80,000	1
5	F	60,000	1

No tuple with DNO=4, Sex='F' because group (set) is empty!

Example: Group By Aggregate (3)

SSN	FName	Other	Sex	DNo	Salary
111-11-1111	John	…	M	4	50,000
242-12-2340	Mary	…	F	5	60,000
222-22-2222	James	…	M	5	80,000
333-33-3333	Jake	…	M	4	60,000

$$\mathcal{F}_{\text{avg(Salary)},\text{count(SSN)}}(\text{EMPLOYEE})$$

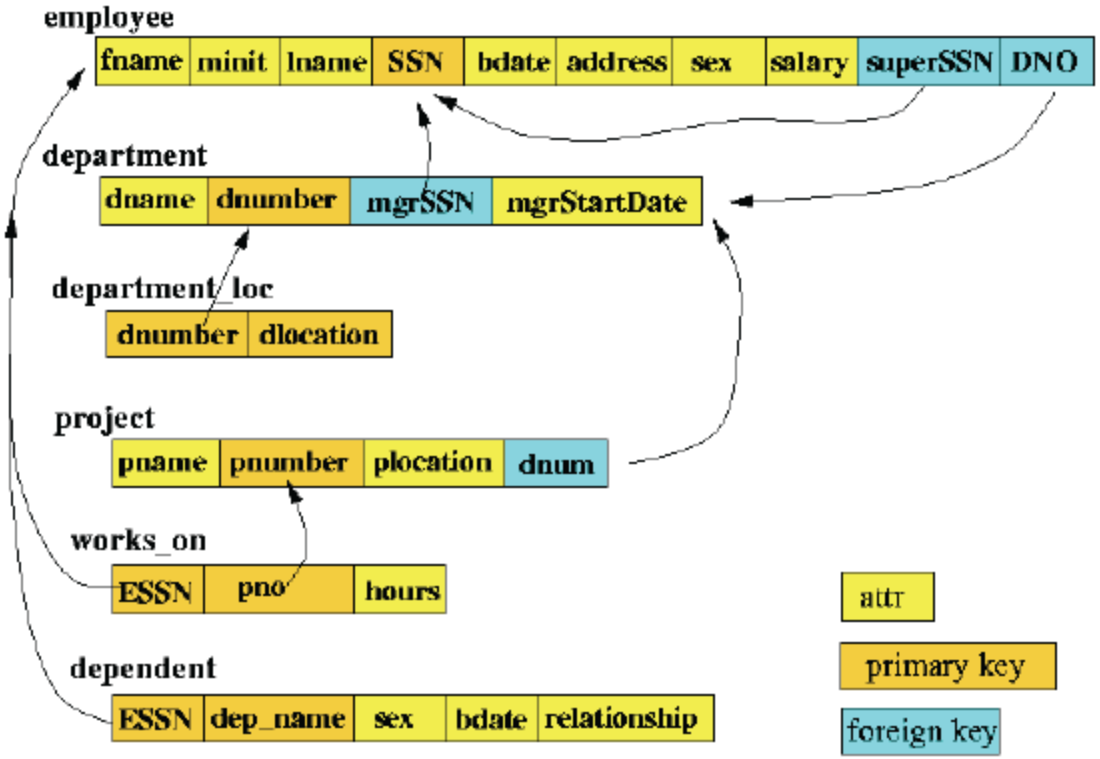
Avg	Count
62,500	4

When no grouping attributes are specified, the set function is applied on ONE group with all the tuples in the relation!

Relational Algebra Operations

Operation	Notation	Purpose
SELECT	$\sigma_{<\text{selection condition}>}(R)$	Selects all tuples that satisfy the selection condition from a relation R
PROJECT	$\pi_{<\text{atttribute list}>}(R)$	New relation with subset of attributes of R and removes duplicate tuples
THETA_JOIN	$R_1 \bowtie_{<\text{join condition}>} R_2$	All combinations of tuples from R_1 and R_2 that satisfy the join condition
EQUIJOIN	$R_1 \bowtie_{<\text{join condition}>} R_2$	Theta join with only equality join comparisons
NATURAL JOIN	$R_1 *_{<\text{join condition}>} R_2$	Equijoin except join attributes of R_2 are not included in the resulting relation
UNION	$R_1 \cup R_2$	Relation that includes all tuples in R_1 or R_2
INTERSECTION	$R_1 \cap R_2$	Relation that includes all tuples in both R_1 and R_2
DIFFERENCE	$R_1 - R_2$	Relation that includes all tuples in R_1 that are not in R_2
CARTESIAN PRODUCT	$R_1 \times R_2$	Relation with attributes of R_1 and R_2 and includes tuples with all possible combinations of tuples of R_1 and R_2
DIVISION	$R_1(Z) \div R_2(Y)$	Relation that includes all tuples t[X] in $R_1(Z)$ that appear in R_1 in combination with every tuple from $R_2(Y)$ where $Z = X \cup Y$

Example: Company Database



Example: RA Queries (1)

Find the name and address of all employees who work in the Research department

Example: RA Queries (2)

Find fname and lname of employees who earn more than ‘John Smith’

Example: RA Queries (3)

Find fname and lname of employees who have 2 or more dependents

Example: RA Queries (4)

Find fname and lname of employees who have the most number of dependents

Example: RA Queries (5)

Retrieve the names of employees who have no dependents

Example: RA Queries (6)

List the names of managers who have at least one dependent

Example: RA Queries (7)

Find fname and lname of employees who work on more projects than ‘John Smith’

Example: RA Queries (8)

For each department, show the department name, number of employees, minimum employee salary and maximum employee salary

Example: RA Queries (9)

Find fname and lname of all employees who work on 2 or more projects controlled by the Research department

Example: RA Queries (10)

Find fname and lname of all employees who work on all projects controlled by the Research department

Example: RA Queries (11)

Find fname and lname of all employees who do not work on any projects controlled by the Research department

Example: RA Queries (12)

Find fname and lname of all employees that only work on projects controlled by the Research department