Lecture 7

ALIASES

- **In SQL**, we can use the **same name for two or more attributes** as long as the attributes are in ***different relations***

A query that refers to two attributes with the same name must *prefix* the relation name to the attribute name

Example:

EMPLOYEE.DNO, DEPARTMENT.DNO

ALIASES

For each employee, retrieve the employee's name, and the name of his or her immediate supervisor.

Using **AS** keyword to specify aliases

```
SELECT  E.FNAME, E.LNAME, S.FNAME, S.LNAME
FROM    EMPLOYEE AS E, EMPLOYEE AS S
WHERE   E.SUPERSSN=S.SSN
```

```
SELECT  E.FNAME, E.LNAME, S.FNAME, S.LNAME
FROM    EMPLOYEE E S
WHERE   E.SUPERSSN=S.SSN
```

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address
John	B	Smith	123456789	1965-01-09	731 Fondren, Ho
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houst
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Sp
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bella
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, H
Joyce	A	English	453453453	1972-07-31	5631 Rice, Hous
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Hou
James	E	Borg	888665555	1937-11-10	450 Stone, Hous

E.Fname	E.Lname	S.Fname	S.Lname
John	Smith	Franklin	Wong
Franklin	Wong	James	Borg
Alicia	Zelaya	Jennifer	Wallace
Jennifer	Wallace	James	Borg
Ramesh	Narayan	Franklin	Wong
Joyce	English	Franklin	Wong
Ahmad	Jabbar	Jennifer	Wallace

ARITHMETIC OPERATIONS

Arithmetic operators '+', '-'. '*', and '/') can be applied to numeric values in an SQL query result

Give all employees who work on the 'ProductX' project a 10% raise.

```
SELECT    FNAME, LNAME, 1.1*SALARY
FROM      (WORKS_ON JOIN PROJECT ON PNO=PNUMBER)
          JOIN EMPLOYEE ON ESSN=SSN
WHERE     PNAME='ProductX'
```

EMPLOYEE

FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO

PROJECT

PNAME	PNUMBER	PLOCATION	DNUM

WORKS_ON

ESSN	PNO	HOURS

ORDER BY

The **ORDER BY** clause **sort** the ***tuples*** in a query result

Retrieve a list of employees and the projects each works in, ordered by the employee's department number

```
SELECT    DNO, LNAME, FNAME, PNO
FROM      EMPLOYEE JOIN WORKS_ON ON SSN=ESSN
ORDER BY  DNO
```

EMPLOYEE

FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO

DEPARTMENT

DNAME	DNUMBER	MGRSSN	MGRSTARTDATE

PROJECT

PNAME	PNUMBER	PLOCATION	DNUM

DEPT_LOCATIONS

DNUMBER	DLOCATION

WORKS_ON

ESSN	PNO	HOURS

ORDER BY

The **ORDER BY** clause **sort** the ***tuples*** in a query result

Retrieve a list of employees and the projects each works in, ordered by the employee's department number

```
SELECT    DNO, LNAME, FNAME, PNO
FROM      EMPLOYEE JOIN WORKS_ON ON SSN=ESSN
ORDER BY  DNO
```

The default order is in ***ascending order*** of values

We can specify the keyword **DESC** if we want a descending order

- **ORDER BY** Dname **DESC,** Lname **ASC**

ORDER BY

Retrieve a list of Male employees and the projects each works in, ordered by the employee's department **name**, **and within each department ordered alphabetically by employee last name, then first name.**

```
SELECT    DNAME, LNAME, FNAME, PNAME
FROM      ((DEPARTMENT JOIN EMPLOYEE ON DNUMBER=DNO )
               JOIN WORKS_ON ON SSN=ESSN )
               JOIN PROJECT ON PNO=PNUMBER)
WHERE     SEX = 'MALE'
ORDER BY  DNAME, LNAME, FNAME
```

SQL QUERIES

- Retrieve the names of all employees who do not have supervisors.
-

```
SELECT    FNAME, LNAME
FROM      EMPLOYEE
WHERE     SUPERSSN IS NULL
```

- **Note:** **If a join condition is specified, tuples with NULL values for the join attributes are not included in the result**

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

SUBSTRING COMPARISON

LIKE operator is used to compare partial strings

Two reserved characters are used:

- '%' (or '*' in some implementations) replaces an arbitrary number of characters, and
- '_' replaces a single arbitrary character

Retrieve all employees whose address is in Houston, Texas.

- **SELECT FNAME, LNAME**
  FROM EMPLOYEE
  WHERE ADDRESS LIKE '%Houston,TX%'

SUBSTRING COMPARISON

Retrieve all employees who were born during the 1950s.

- **SELECT** **FNAME, LNAME**
  FROM **EMPLOYEE**
  WHERE **BDATE LIKE** '195_____',.

LIKE operator allows us to get around the fact that each value is considered atomic and indivisible

Hence, in SQL, character string attribute values are not atomic

AGGREGATE FUNCTIONS

- Include **COUNT**, **SUM**, **MAX**, **MIN**, and **AVG**
- Find the maximum salary, the minimum salary, and the average salary among all employees.

```
SELECT   MAX(SALARY), MIN(SALARY), AVG(SALARY)
FROM     EMPLOYEE
```

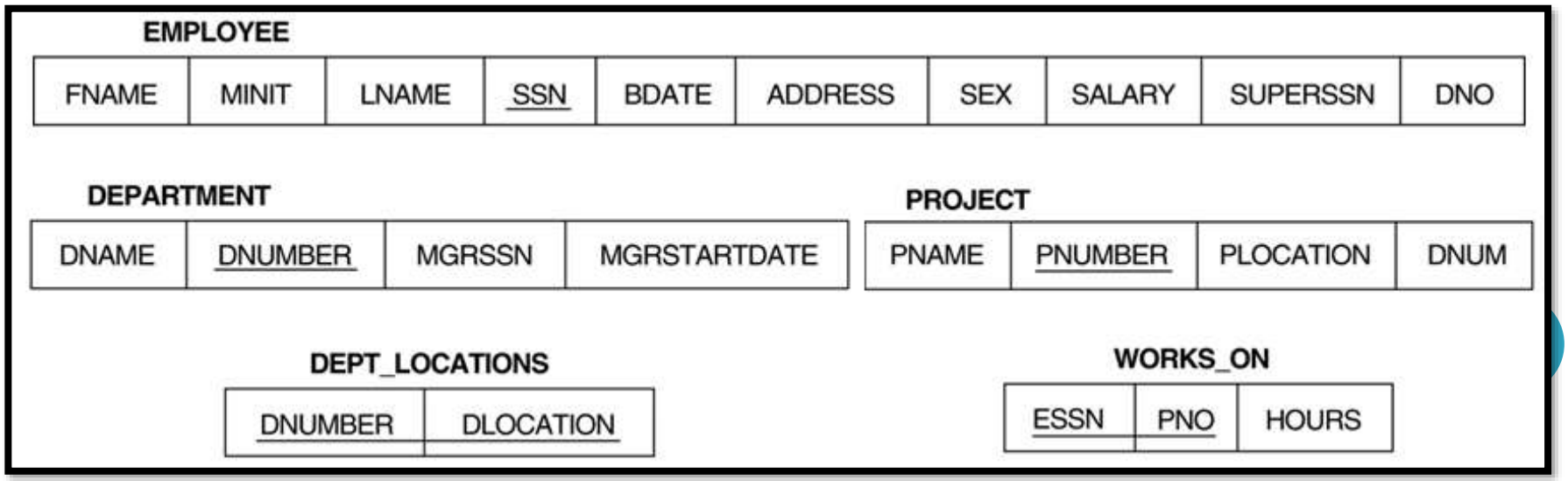
$\mathcal{F}_{\text{SUM Salary, AVERAGE Salary, MIN Salary}}$ (EMPLOYEE)

Some SQL implementations *may not allow more than one function* in the SELECT-clause

AGGREGATE FUNCTIONS

Retrieve the no. of employees in the ‘Administration' department

```
SELECT   COUNT (*)
FROM     EMPLOYEE JOIN DEPARTMENT ON DNO=DNUMBER
WHERE    DNAME=‘Administration’
```



Aggregate Functions $\mathcal{F}$

- $\mathcal{F}_{\text{MAX Salary}}$ (EMPLOYEE)
- $\mathcal{F}_{\text{MIN Salary}}$ (EMPLOYEE)
- $\mathcal{F}_{\text{SUM Salary, AVERAGE Salary}}$ (EMPLOYEE)
- $\mathcal{F}_{\text{COUNT SSN}}$ (EMPLOYEE)

COUNT (*) returns the no. of rows in the result of the query ***(it counts without removing duplicates)***

NULL values are **discarded** when aggregate functions are applied to a particular column (attribute).

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

AGGREGATE EXAMPLE

Count the number of distinct salary values in the database

```
SELECT  COUNT (DISTINCT Salary)
FROM    EMPLOYEE
```

NULL values are **discarded** when aggregate functions are applied to a particular attribute.

Aggregate functions are allowed only in the SELECT and the HAVING clause of a SQL statement.

GROUPING

GROUP BY-clause specifies the grouping attributes

For each department, retrieve the department number, the number of employees in the department, and their average salary.

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

GROUPING

For each department, retrieve the DNO, the number of employees, and their average salary.

```
SELECT      DNO, COUNT (*), AVG (SALARY)
FROM        EMPLOYEE
GROUP BY    DNO
```

$$\text{DNO}\ \mathcal{F}\ \text{COUNT SSN, AVERAGE Salary}\ \text{EMPLOYEE}$$

Fname	Minit	Lname	Ssn	...	Salary	Super_ssn	Dno
John	B	Smith	123456789		30000	333445555	5
Franklin	T	Wong	333445555		40000	888665555	5
Ramesh	K	Narayan	666884444		38000	333445555	5
Joyce	A	English	453453453	...	25000	333445555	5
Alicia	J	Zelaya	999887777		25000	987654321	4
Jennifer	S	Wallace	987654321		43000	888665555	4
Ahmad	V	Jabbar	987987987		25000	987654321	4
James	E	Borg	888665555		55000	NULL	1

Grouping EMPLOYEE tuples by the value of Dno

Dno	Count_ssn	Average_salary
5	4	33250
4	3	31000
1	1	55000

Grouping with Aggregation

$DNO\ \mathcal{F}_{COUNT\ SSN,\ AVERAGE\ Salary}$ EMPLOYEE

Dno	Count_ssn	Average_salary
5	4	33250
4	3	31000
1	1	55000

$\mathcal{F}_{COUNT\ SSN,\ AVERAGE\ Salary}$ EMPLOYEE

Count_ssn	Average_salary
8	35125

ρ_R(Dno, No_of_employees, Average_sal) ($_{DNO}\mathcal{F}_{COUNT\ SSN,\ AVERAGE\ Salary}$ EMPLOYEE)

R

Dno	No_of_employees	Average_sal
5	4	33250
4	3	31000
1	1	55000

GROUPING

For each project, retrieve the **project number**, **project name**, and the **number of employees** who work on that project.

EMPLOYEE

FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO

DEPARTMENT

DNAME	DNUMBER	MGRSSN	MGRSTARTDATE

PROJECT

PNAME	PNUMBER	PLOCATION	DNUM

DEPT_LOCATIONS

DNUMBER	DLOCATION

WORKS_ON

ESSN	PNO	HOURS

GROUPING

For each project, retrieve the project number, project name, and the number of employees who work on that project.

```
SELECT      PNUMBER, PNAME, COUNT (*)
FROM        PROJECT, WORKS_ON
WHERE       PNUMBER=PNO
GROUP BY    PNUMBER, PNAME
```

	PNUMBER	PNAME	count
1	1	ProductX	2
2	2	ProductY	3
3	3	ProductZ	2
4	10	Compu...	3
5	20	Reorga...	3
6	30	Newbe...	3

Grouping & Aggregate functions are applied *after* the joining two relations

Select clause can only include the grouping attributes and aggregate functions

Example: Retrieve the names of all employees with two or more dependents

$$T1(Ssn, No\_of\_dependents) \leftarrow Essn\ \mathcal{F}_{COUNT\ Dependent\_name}\ DEPENDENT$$

$$T2 \leftarrow \sigma_{No\_of\_dependents > 1}(T1)$$

$$RESULT \leftarrow \pi_{LNAME, FNAME}\ (T2 * EMPLOYEE)$$

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

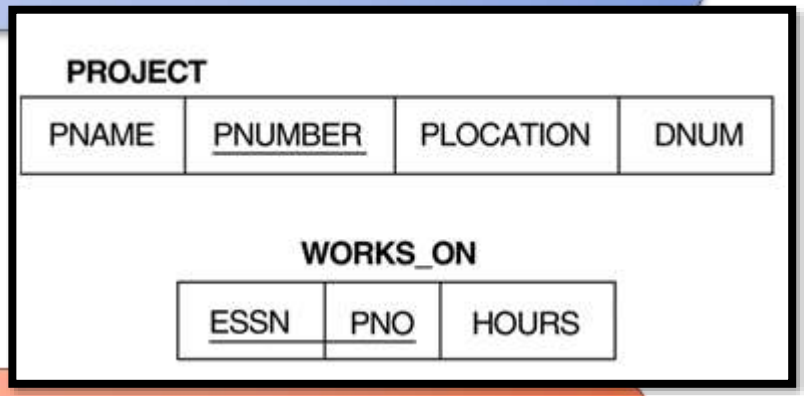
EMPLOYEE

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

HAVING-CLAUSE

HAVING-clause specify a selection condition on groups

For each project *on which more than two employees work* , retrieve the project number, name, and the number of employees who work on that project.




```
SELECT      PNUMBER, PNAME, COUNT (*)
FROM        PROJECT JOIN WORKS_ON ON
                    PNUMBER=PNO
GROUP BY    PNUMBER, PNAME
HAVING      COUNT (*) > 2
```

HAVING-CLAUSE

Pname	Pnumber	...	Essn	Pno	Hours
ProductX	1		123456789	1	32.5
ProductX	1		453453453	1	20.0
ProductY	2		123456789	2	7.5
ProductY	2		453453453	2	20.0
ProductY	2		333445555	2	10.0
ProductZ	3		666884444	3	40.0
ProductZ	3		333445555	3	10.0
Computerization	10	...	333445555	10	10.0
Computerization	10		999887777	10	10.0
Computerization	10		987987987	10	35.0
Reorganization	20		333445555	20	10.0
Reorganization	20		987654321	20	15.0
Reorganization	20		888665555	20	NULL
Newbenefits	30		987987987	30	5.0
Newbenefits	30		987654321	30	20.0
Newbenefits	30		999887777	30	30.0

After applying the WHERE clause but before applying HAVING

These groups are not selected by the HAVING condition of Q26.

PNUMBER	PNAME	count
2	ProductY	3
10	Computerization	3
20	Reorganization	3
30	Newbenefits	3

HAVING-CLAUSE

Pname	Pnumber	...	Essn	Pno	Hours
ProductY	2		123456789	2	7.5
ProductY	2		453453453	2	20.0
ProductY	2		333445555	2	10.0
Computerization	10		333445555	10	10.0
Computerization	10	...	999887777	10	10.0
Computerization	10		987987987	10	35.0
Reorganization	20		333445555	20	10.0
Reorganization	20		987654321	20	15.0
Reorganization	20		888665555	20	NULL
Newbenefits	30		987987987	30	5.0
Newbenefits	30		987654321	30	20.0
Newbenefits	30		999887777	30	30.0

After applying the HAVING clause condition

Pname	Count (*)
ProductY	3
Computerization	3
Reorganization	3
Newbenefits	3

Result of Q26
(Pnumber not shown)

General form of Grouping and Aggregation

```
SELECT    S
FROM      R1,…,Rn
WHERE     C1
GROUP BY  a1,…,ak
HAVING    C2
```

Evaluation steps:

- Evaluate FROM-WHERE, apply condition C1
- Group by the attributes $a_1,\ldots,a_k$
- Apply condition C2 to each group (may have aggregates)
- Compute aggregates in S and return the result

Example

- List the employees name and the department name that they manage.
- Temp ← (Employee $\bowtie_{Ssn=Mgr\_Ssn}$ Department)
- Result ← $\pi_{Fname, Minit, Lname, Dname}$(Temp)

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

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

Left Outer Join

- List the employees name and the department name that they manage. **If they don't manage one, then indicate this with a null value.**
- Temp ← (Employee $\overset{\bowtie}{}_{Ssn=Mgr\_Ssn}$ Department)
- Result ← $\pi_{Fname, Minit, Lname, Dname}$(Temp)

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

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

Left Outer Join

- List the employees name and the department name that they manage. **If they don't manage one, then indicate this with a null value.**
- Temp ← (Employee $\overset{\text{Ssn=Mgr\_Ssn}}{\bowtie}$ Department)
- Result ← $\pi_{\text{Fname, Minit, Lname, Dname}}$(Temp)

RESULT

Fname	Minit	Lname	Dname
John	B	Smith	NULL
Franklin	T	Wong	Research
Alicia	J	Zelaya	NULL
Jennifer	S	Wallace	Administration
Ramesh	K	Narayan	NULL
Joyce	A	English	NULL
Ahmad	V	Jabbar	NULL
James	E	Borg	Headquarters

Right Outer Join

- List the employees name and the department name that they manage. If they don't manage one, then indicate this with a null value.
- Temp ← (**Department** $\overset{}{\bowtie}_{Mgr\_Ssn= Ssn}$ **Employee**)
- Result ← $\pi_{Fname, Minit, Lname, Dname}$(Temp)

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

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