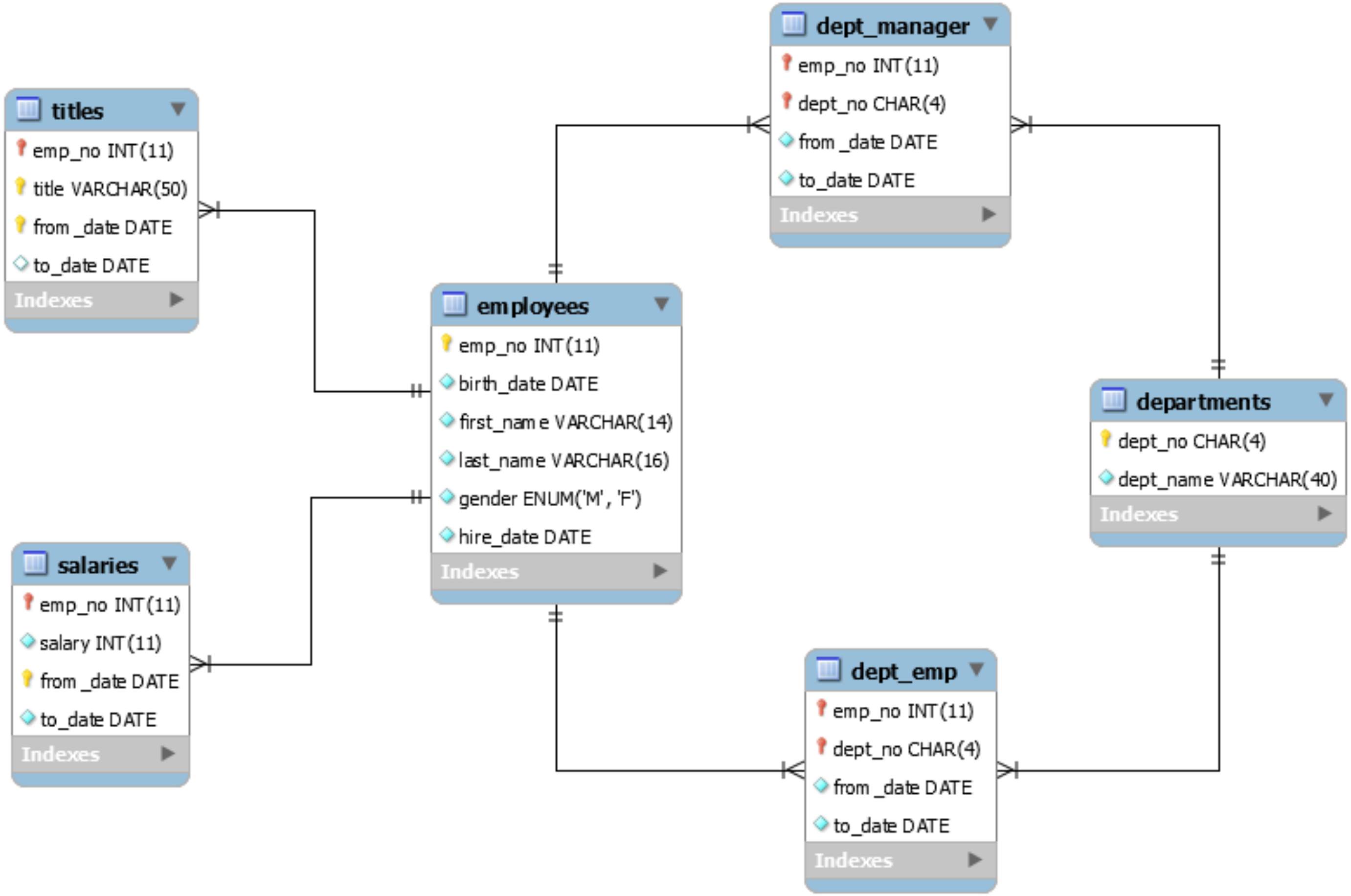
TÉCNICO
LISBOA

Data Analysis and Integration

A review of SQL

Introduction



titles
- emp_no INT(11)
- title VARCHAR(50)
- from _date DATE
- to_date DATE
- Indexes

dept_manager
- emp_no INT(11)
- dept_no CHAR(4)
- from _date DATE
- to_date DATE
- Indexes

employees
- emp_no INT(11)
- birth_date DATE
- first_nam e VARCHAR(14)
- last_name VARCHAR(16)
- gender ENUM('M', 'F')
- hire_date DATE
- Indexes

departments
- dept_no CHAR(4)
- dept_name VARCHAR(40)
- Indexes

salaries
- emp_no INT(11)
- salary INT(11)
- from _date DATE
- to_date DATE
- Indexes

dept_emp
- emp_no INT(11)
- dept_no CHAR(4)
- from _date DATE
- to_date DATE
- Indexes

Introduction

Introduction

The employees table

select * **from** employees **limit** 10;

```
+--------+------------+------------+-------------+--------+------------+
| emp_no | birth_date | first_name | last_name   | gender | hire_date  |
+--------+------------+------------+-------------+--------+------------+
|  10721 | 1964-07-12 | Bernd      | Redmiles    | F      | 1995-03-17 |
|  11260 | 1963-06-11 | Ingemar    | Schade      | M      | 1993-09-29 |
|  11371 | 1956-05-17 | Tadahiko   | Masamoto    | M      | 1988-12-25 |
|  11693 | 1952-02-15 | Hideo      | Coorg       | M      | 1989-03-13 |
|  13816 | 1964-03-09 | Miquel     | Maksimenko  | F      | 1992-09-17 |
|  14007 | 1961-12-02 | Shiv       | Jervis      | M      | 1987-02-03 |
|  14083 | 1959-05-27 | Sashi      | Figueira    | M      | 1986-04-13 |
|  14131 | 1960-02-17 | Kirk       | Thorensen   | M      | 1993-02-21 |
|  14791 | 1952-12-12 | Magy       | Garrabrants | F      | 1992-02-17 |
|  17698 | 1959-12-03 | Kazuhito   | Larfeldt    | M      | 1988-08-07 |
+--------+------------+------------+-------------+--------+------------+
                        10 rows in set (0.00 sec)
```

Introduction

The departments table

select * **from** departments;

```
+---------+--------------------+
| dept_no | dept_name          |
+---------+--------------------+
| d009    | Customer Service   |
| d005    | Development        |
| d002    | Finance            |
| d003    | Human Resources    |
| d001    | Marketing          |
| d004    | Production         |
| d006    | Quality Management |
| d008    | Research           |
| d007    | Sales              |
+---------+--------------------+
    9 rows in set (0.00 sec)
```

Introduction

The dept_emp table

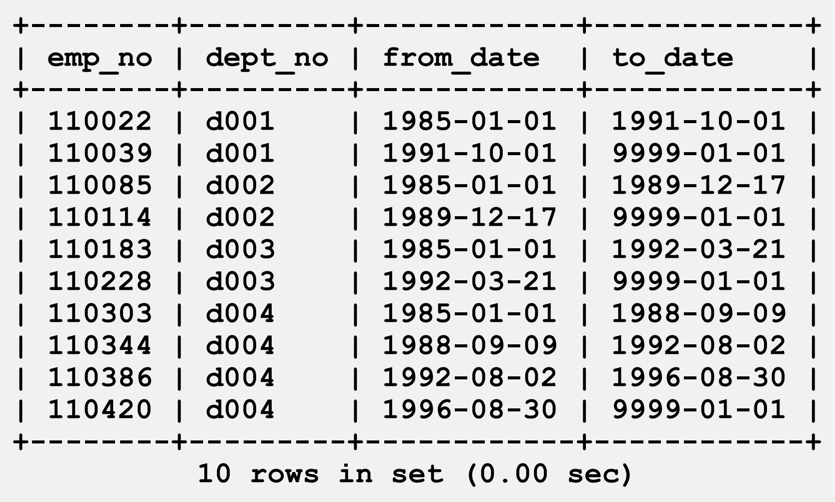
select * **from** dept_emp **limit** 10;

```
+--------+---------+------------+------------+
| emp_no | dept_no | from_date  | to_date    |
+--------+---------+------------+------------+
|  10721 | d009    | 1999-07-15 | 9999-01-01 |
|  11260 | d009    | 1996-05-16 | 9999-01-01 |
|  11371 | d005    | 1988-12-25 | 9999-01-01 |
|  11693 | d005    | 1991-08-24 | 9999-01-01 |
|  13816 | d005    | 1992-09-17 | 9999-01-01 |
|  14007 | d002    | 1988-10-17 | 9999-01-01 |
|  14083 | d004    | 1986-04-13 | 9999-01-01 |
|  14131 | d004    | 1993-05-11 | 1996-05-28 |
|  14791 | d005    | 1997-11-16 | 9999-01-01 |
|  17698 | d005    | 1988-08-07 | 9999-01-01 |
+--------+---------+------------+------------+
           10 rows in set (0.00 sec)
```

Introduction

The dept_manager table

select * **from** dept_manager **limit** 10;



```
+--------+---------+------------+------------+
| emp_no | dept_no | from_date  | to_date    |
+--------+---------+------------+------------+
| 110022 | d001    | 1985-01-01 | 1991-10-01 |
| 110039 | d001    | 1991-10-01 | 9999-01-01 |
| 110085 | d002    | 1985-01-01 | 1989-12-17 |
| 110114 | d002    | 1989-12-17 | 9999-01-01 |
| 110183 | d003    | 1985-01-01 | 1992-03-21 |
| 110228 | d003    | 1992-03-21 | 9999-01-01 |
| 110303 | d004    | 1985-01-01 | 1988-09-09 |
| 110344 | d004    | 1988-09-09 | 1992-08-02 |
| 110386 | d004    | 1992-08-02 | 1996-08-30 |
| 110420 | d004    | 1996-08-30 | 9999-01-01 |
+--------+---------+------------+------------+
           10 rows in set (0.00 sec)
```

Introduction

The titles table

select * **from** titles **limit** 10;

```
+--------+--------------------+------------+------------+
| emp_no | title              | from_date  | to_date    |
+--------+--------------------+------------+------------+
|  10721 | Staff              | 1999-07-15 | 9999-01-01 |
|  11260 | Staff              | 1996-05-16 | 9999-01-01 |
|  11371 | Assistant Engineer | 1988-12-25 | 1993-12-25 |
|  11371 | Engineer           | 1993-12-25 | 1998-12-25 |
|  11371 | Senior Engineer    | 1998-12-25 | 9999-01-01 |
|  11693 | Senior Engineer    | 1991-08-24 | 9999-01-01 |
|  13816 | Engineer           | 1992-09-17 | 1999-09-18 |
|  13816 | Senior Engineer    | 1999-09-18 | 9999-01-01 |
|  14007 | Senior Staff       | 1995-10-18 | 9999-01-01 |
|  14007 | Staff              | 1988-10-17 | 1995-10-18 |
+--------+--------------------+------------+------------+
             10 rows in set (0.00 sec)
```

Introduction

The salaries table

select * **from** salaries **limit** 10;

```
+--------+--------+------------+------------+
| emp_no | salary | from_date  | to_date    |
+--------+--------+------------+------------+
|  10721 |  40000 | 1999-07-15 | 2000-07-14 |
|  10721 |  40477 | 2000-07-14 | 2001-07-14 |
|  10721 |  41015 | 2001-07-14 | 2002-07-14 |
|  10721 |  44812 | 2002-07-14 | 9999-01-01 |
|  11260 |  40000 | 1996-05-16 | 1997-05-16 |
|  11260 |  41740 | 1997-05-16 | 1998-05-16 |
|  11260 |  45525 | 1998-05-16 | 1999-05-16 |
|  11260 |  45850 | 1999-05-16 | 2000-05-15 |
|  11260 |  47692 | 2000-05-15 | 2001-05-15 |
|  11260 |  48889 | 2001-05-15 | 2002-05-15 |
+--------+--------+------------+------------+
          10 rows in set (0.00 sec)
```

Simple SQL

Simple queries

Queries with selection (conditions)

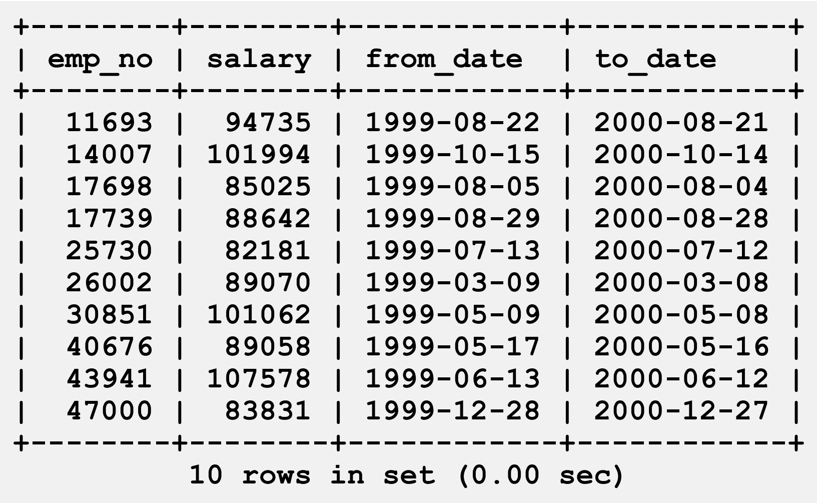
```
select *
from salaries
where salary > 80000
limit 10;
```

```
+--------+--------+------------+------------+
| emp_no | salary | from_date  | to_date    |
+--------+--------+------------+------------+
|  11371 |  81461 | 2001-12-22 | 9999-01-01 |
|  11693 |  80506 | 1994-08-23 | 1995-08-23 |
|  11693 |  83059 | 1995-08-23 | 1996-08-22 |
|  11693 |  86434 | 1996-08-22 | 1997-08-22 |
|  11693 |  86865 | 1997-08-22 | 1998-08-22 |
|  11693 |  91258 | 1998-08-22 | 1999-08-22 |
|  11693 |  94735 | 1999-08-22 | 2000-08-21 |
|  11693 |  97681 | 2000-08-21 | 2001-08-21 |
|  11693 | 101179 | 2001-08-21 | 9999-01-01 |
|  14007 |  82180 | 1990-10-17 | 1991-10-17 |
+--------+--------+------------+------------+
          10 rows in set (0.00 sec)
```

Simple queries

Queries with selection (conditions)

```
select *
from salaries
where salary > 80000
    and from_date <= '2000-01-01'
     and to_date >= '2000-01-01'
limit 10;
```

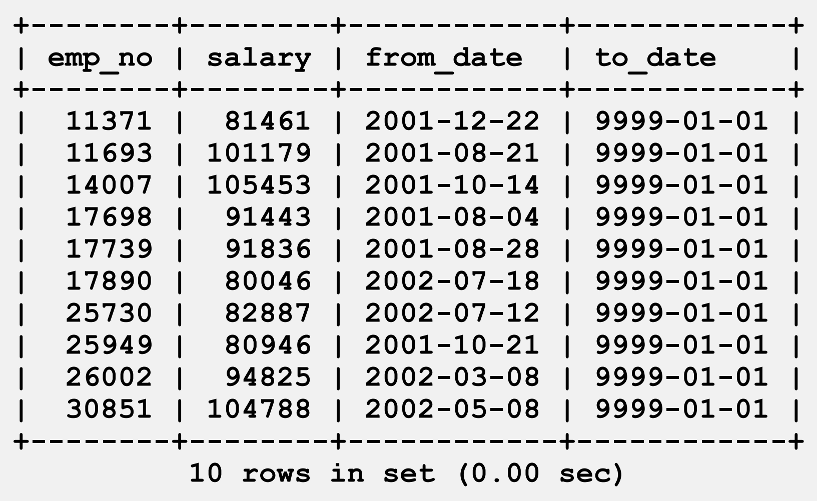


```
+--------+--------+------------+------------+
| emp_no | salary | from_date  | to_date    |
+--------+--------+------------+------------+
|  11693 |  94735 | 1999-08-22 | 2000-08-21 |
|  14007 | 101994 | 1999-10-15 | 2000-10-14 |
|  17698 |  85025 | 1999-08-05 | 2000-08-04 |
|  17739 |  88642 | 1999-08-29 | 2000-08-28 |
|  25730 |  82181 | 1999-07-13 | 2000-07-12 |
|  26002 |  89070 | 1999-03-09 | 2000-03-08 |
|  30851 | 101062 | 1999-05-09 | 2000-05-08 |
|  40676 |  89058 | 1999-05-17 | 2000-05-16 |
|  43941 | 107578 | 1999-06-13 | 2000-06-12 |
|  47000 |  83831 | 1999-12-28 | 2000-12-27 |
+--------+--------+------------+------------+
          10 rows in set (0.00 sec)
```

Simple queries

Queries with selection (conditions)

```
select *
from salaries
where salary > 80000
    and from_date <= current_date
     and to_date >= current_date
limit 10;
```



```
+--------+--------+------------+------------+
| emp_no | salary | from_date  | to_date    |
+--------+--------+------------+------------+
|  11371 |  81461 | 2001-12-22 | 9999-01-01 |
|  11693 | 101179 | 2001-08-21 | 9999-01-01 |
|  14007 | 105453 | 2001-10-14 | 9999-01-01 |
|  17698 |  91443 | 2001-08-04 | 9999-01-01 |
|  17739 |  91836 | 2001-08-28 | 9999-01-01 |
|  17890 |  80046 | 2002-07-18 | 9999-01-01 |
|  25730 |  82887 | 2002-07-12 | 9999-01-01 |
|  25949 |  80946 | 2001-10-21 | 9999-01-01 |
|  26002 |  94825 | 2002-03-08 | 9999-01-01 |
|  30851 | 104788 | 2002-05-08 | 9999-01-01 |
+--------+--------+------------+------------+
          10 rows in set (0.00 sec)
```

Simple queries

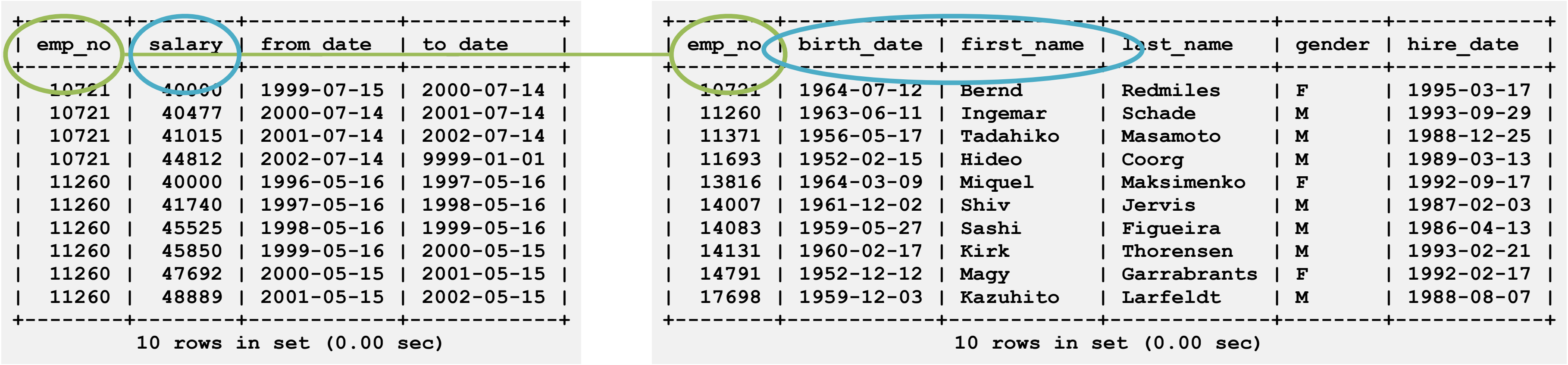
Queries with projection

```
select emp_no, salary
from salaries
where salary > 80000
   and from_date <= current_date
    and to_date >= current_date
limit 10;
```

```
+--------+---------+
| emp_no | salary  |
+--------+---------+
|  11371 |   81461 |
|  11693 |  101179 |
|  14007 |  105453 |
|  17698 |   91443 |
|  17739 |   91836 |
|  17890 |   80046 |
|  25730 |   82887 |
|  25949 |   80946 |
|  26002 |   94825 |
|  30851 |  104788 |
+--------+---------+
10 rows in set (0.00 sec)
```

Simple queries

Joining tables: salaries and employees



salaries

```
+--------+--------+------------+------------+
| emp_no | salary | from_date  | to_date    |
+--------+--------+------------+------------+
|  10721 |  40000 | 1999-07-15 | 2000-07-14 |
|  10721 |  40477 | 2000-07-14 | 2001-07-14 |
|  10721 |  41015 | 2001-07-14 | 2002-07-14 |
|  10721 |  44812 | 2002-07-14 | 9999-01-01 |
|  11260 |  40000 | 1996-05-16 | 1997-05-16 |
|  11260 |  41740 | 1997-05-16 | 1998-05-16 |
|  11260 |  45525 | 1998-05-16 | 1999-05-16 |
|  11260 |  45850 | 1999-05-16 | 2000-05-15 |
|  11260 |  47692 | 2000-05-15 | 2001-05-15 |
|  11260 |  48889 | 2001-05-15 | 2002-05-15 |
+--------+--------+------------+------------+
          10 rows in set (0.00 sec)
```

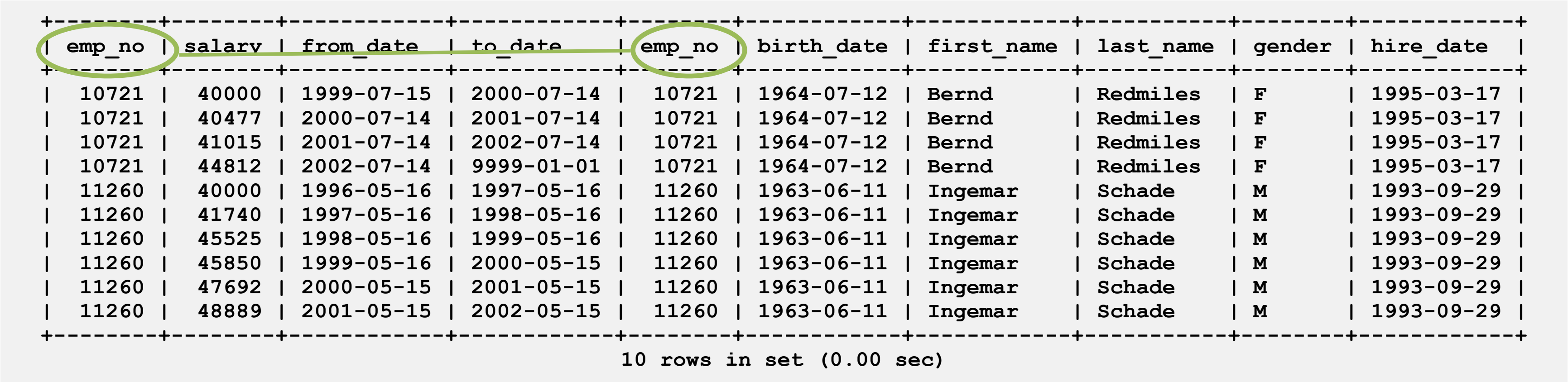
employees

```
+--------+------------+------------+-------------+--------+------------+
| emp_no | birth_date | first_name | last_name   | gender | hire_date  |
+--------+------------+------------+-------------+--------+------------+
|  10721 | 1964-07-12 | Bernd      | Redmiles    | F      | 1995-03-17 |
|  11260 | 1963-06-11 | Ingemar    | Schade      | M      | 1993-09-29 |
|  11371 | 1956-05-17 | Tadahiko   | Masamoto    | M      | 1988-12-25 |
|  11693 | 1952-02-15 | Hideo      | Coorg       | M      | 1989-03-13 |
|  13816 | 1964-03-09 | Miquel     | Maksimenko  | F      | 1992-09-17 |
|  14007 | 1961-12-02 | Shiv       | Jervis      | M      | 1987-02-03 |
|  14083 | 1959-05-27 | Sashi      | Figueira    | M      | 1986-04-13 |
|  14131 | 1960-02-17 | Kirk       | Thorensen   | M      | 1993-02-21 |
|  14791 | 1952-12-12 | Magy       | Garrabrants | F      | 1992-02-17 |
|  17698 | 1959-12-03 | Kazuhito   | Larfeldt    | M      | 1988-08-07 |
+--------+------------+------------+-------------+--------+------------+
                       10 rows in set (0.00 sec)
```

Simple queries

Joining tables: salaries and employees

select *
from salaries **as** a, employees **as** b
where a.emp_no = b.emp_no
limit 10;

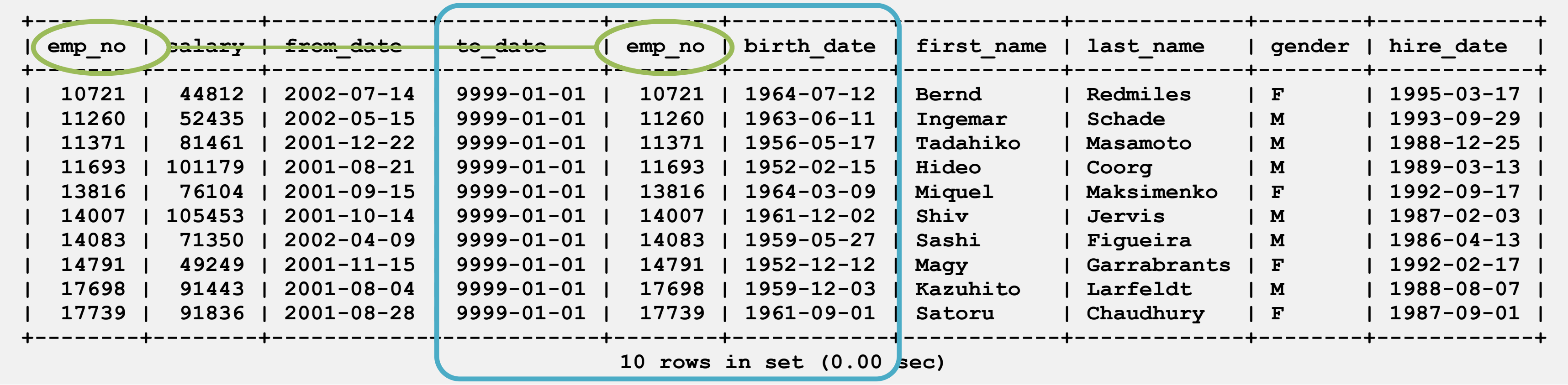


```
+--------+--------+------------+------------+--------+------------+------------+-----------+--------+------------+
| emp_no | salary | from_date  | to_date    | emp_no | birth_date | first_name | last_name | gender | hire_date  |
+--------+--------+------------+------------+--------+------------+------------+-----------+--------+------------+
|  10721 |  40000 | 1999-07-15 | 2000-07-14 |  10721 | 1964-07-12 | Bernd      | Redmiles  | F      | 1995-03-17 |
|  10721 |  40477 | 2000-07-14 | 2001-07-14 |  10721 | 1964-07-12 | Bernd      | Redmiles  | F      | 1995-03-17 |
|  10721 |  41015 | 2001-07-14 | 2002-07-14 |  10721 | 1964-07-12 | Bernd      | Redmiles  | F      | 1995-03-17 |
|  10721 |  44812 | 2002-07-14 | 9999-01-01 |  10721 | 1964-07-12 | Bernd      | Redmiles  | F      | 1995-03-17 |
|  11260 |  40000 | 1996-05-16 | 1997-05-16 |  11260 | 1963-06-11 | Ingemar    | Schade    | M      | 1993-09-29 |
|  11260 |  41740 | 1997-05-16 | 1998-05-16 |  11260 | 1963-06-11 | Ingemar    | Schade    | M      | 1993-09-29 |
|  11260 |  45525 | 1998-05-16 | 1999-05-16 |  11260 | 1963-06-11 | Ingemar    | Schade    | M      | 1993-09-29 |
|  11260 |  45850 | 1999-05-16 | 2000-05-15 |  11260 | 1963-06-11 | Ingemar    | Schade    | M      | 1993-09-29 |
|  11260 |  47692 | 2000-05-15 | 2001-05-15 |  11260 | 1963-06-11 | Ingemar    | Schade    | M      | 1993-09-29 |
|  11260 |  48889 | 2001-05-15 | 2002-05-15 |  11260 | 1963-06-11 | Ingemar    | Schade    | M      | 1993-09-29 |
+--------+--------+------------+------------+--------+------------+------------+-----------+--------+------------+
10 rows in set (0.00 sec)
```

Simple queries

Joining tables: salaries and employees

```
select *
from salaries as a, employees as b
where a.emp_no = b.emp_no
    and a.from_date <= current_date
     and a.to_date >= current_date
limit 10;
```



```
+--------+--------+------------+------------+--------+------------+------------+--------------+--------+------------+
| emp_no | salary | from_date  | to_date    | emp_no | birth_date | first_name | last_name    | gender | hire_date  |
+--------+--------+------------+------------+--------+------------+------------+--------------+--------+------------+
|  10721 |  44812 | 2002-07-14 | 9999-01-01 |  10721 | 1964-07-12 | Bernd      | Redmiles     | F      | 1995-03-17 |
|  11260 |  52435 | 2002-05-15 | 9999-01-01 |  11260 | 1963-06-11 | Ingemar    | Schade       | M      | 1993-09-29 |
|  11371 |  81461 | 2001-12-22 | 9999-01-01 |  11371 | 1956-05-17 | Tadahiko   | Masamoto     | M      | 1988-12-25 |
|  11693 | 101179 | 2001-08-21 | 9999-01-01 |  11693 | 1952-02-15 | Hideo      | Coorg        | M      | 1989-03-13 |
|  13816 |  76104 | 2001-09-15 | 9999-01-01 |  13816 | 1964-03-09 | Miquel     | Maksimenko   | F      | 1992-09-17 |
|  14007 | 105453 | 2001-10-14 | 9999-01-01 |  14007 | 1961-12-02 | Shiv       | Jervis       | M      | 1987-02-03 |
|  14083 |  71350 | 2002-04-09 | 9999-01-01 |  14083 | 1959-05-27 | Sashi      | Figueira     | M      | 1986-04-13 |
|  14791 |  49249 | 2001-11-15 | 9999-01-01 |  14791 | 1952-12-12 | Magy       | Garrabrants  | F      | 1992-02-17 |
|  17698 |  91443 | 2001-08-04 | 9999-01-01 |  17698 | 1959-12-03 | Kazuhito   | Larfeldt     | M      | 1988-08-07 |
|  17739 |  91836 | 2001-08-28 | 9999-01-01 |  17739 | 1961-09-01 | Satoru     | Chaudhury    | F      | 1987-09-01 |
+--------+--------+------------+------------+--------+------------+------------+--------------+--------+------------+
                                    10 rows in set (0.00 sec)
```

Simple queries

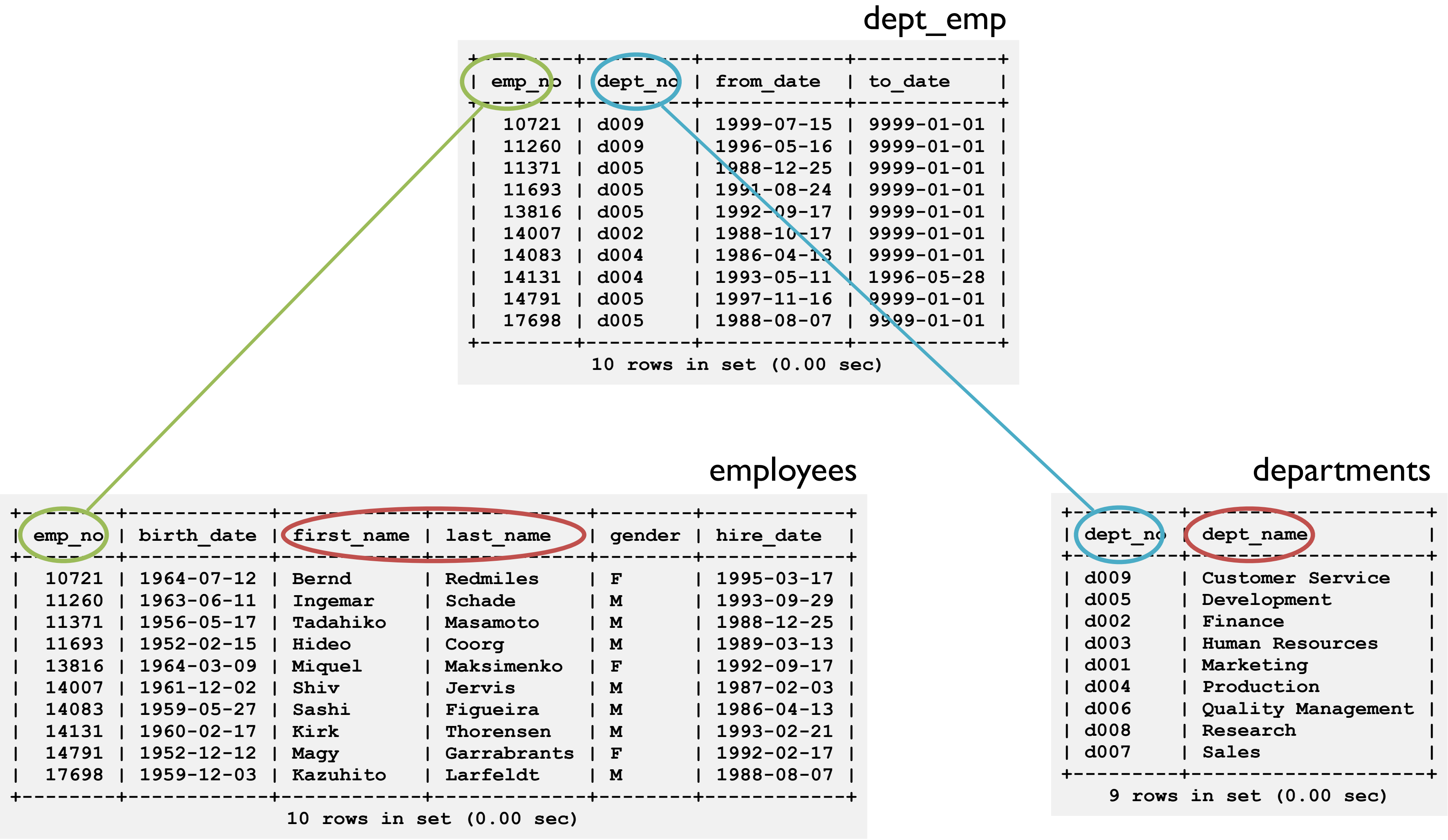
Joining tables: salaries and employees

```
select a.emp_no, a.salary, b.first_name, b.last_name
from salaries as a, employees as b
where a.emp_no = b.emp_no
    and a.from_date <= current_date
     and a.to_date >= current_date
limit 10;
```

```
+--------+--------+------------+-------------+
| emp_no | salary | first_name | last_name   |
+--------+--------+------------+-------------+
|  10721 |  44812 | Bernd      | Redmiles    |
|  11260 |  52435 | Ingemar    | Schade      |
|  11371 |  81461 | Tadahiko   | Masamoto    |
|  11693 | 101179 | Hideo      | Coorg       |
|  13816 |  76104 | Miquel     | Maksimenko  |
|  14007 | 105453 | Shiv       | Jervis      |
|  14083 |  71350 | Sashi      | Figueira    |
|  14791 |  49249 | Magy       | Garrabrants |
|  17698 |  91443 | Kazuhito   | Larfeldt    |
|  17739 |  91836 | Satoru     | Chaudhury   |
+--------+--------+------------+-------------+
           10 rows in set (0.00 sec)
```

Simple queries

Joining tables: employees and departments



dept_emp

```
+--------+---------+------------+------------+
| emp_no | dept_no | from_date  | to_date    |
+--------+---------+------------+------------+
|  10721 | d009    | 1999-07-15 | 9999-01-01 |
|  11260 | d009    | 1996-05-16 | 9999-01-01 |
|  11371 | d005    | 1988-12-25 | 9999-01-01 |
|  11693 | d005    | 1991-08-24 | 9999-01-01 |
|  13816 | d005    | 1992-09-17 | 9999-01-01 |
|  14007 | d002    | 1988-10-17 | 9999-01-01 |
|  14083 | d004    | 1986-04-13 | 9999-01-01 |
|  14131 | d004    | 1993-05-11 | 1996-05-28 |
|  14791 | d005    | 1997-11-16 | 9999-01-01 |
|  17698 | d005    | 1988-08-07 | 9999-01-01 |
+--------+---------+------------+------------+
          10 rows in set (0.00 sec)
```

employees

```
+--------+------------+------------+-------------+--------+------------+
| emp_no | birth_date | first_name | last_name   | gender | hire_date  |
+--------+------------+------------+-------------+--------+------------+
|  10721 | 1964-07-12 | Bernd      | Redmiles    | F      | 1995-03-17 |
|  11260 | 1963-06-11 | Ingemar    | Schade      | M      | 1993-09-29 |
|  11371 | 1956-05-17 | Tadahiko   | Masamoto    | M      | 1988-12-25 |
|  11693 | 1952-02-15 | Hideo      | Coorg       | M      | 1989-03-13 |
|  13816 | 1964-03-09 | Miquel     | Maksimenko  | F      | 1992-09-17 |
|  14007 | 1961-12-02 | Shiv       | Jervis      | M      | 1987-02-03 |
|  14083 | 1959-05-27 | Sashi      | Figueira    | M      | 1986-04-13 |
|  14131 | 1960-02-17 | Kirk       | Thorensen   | M      | 1993-02-21 |
|  14791 | 1952-12-12 | Magy       | Garrabrants | F      | 1992-02-17 |
|  17698 | 1959-12-03 | Kazuhito   | Larfeldt    | M      | 1988-08-07 |
+--------+------------+------------+-------------+--------+------------+
                    10 rows in set (0.00 sec)
```

departments

```
+---------+--------------------+
| dept_no | dept_name          |
+---------+--------------------+
| d009    | Customer Service   |
| d005    | Development        |
| d002    | Finance            |
| d003    | Human Resources    |
| d001    | Marketing          |
| d004    | Production         |
| d006    | Quality Management |
| d008    | Research           |
| d007    | Sales              |
+---------+--------------------+
   9 rows in set (0.00 sec)
```

Simple queries

Joining tables: employees and departments

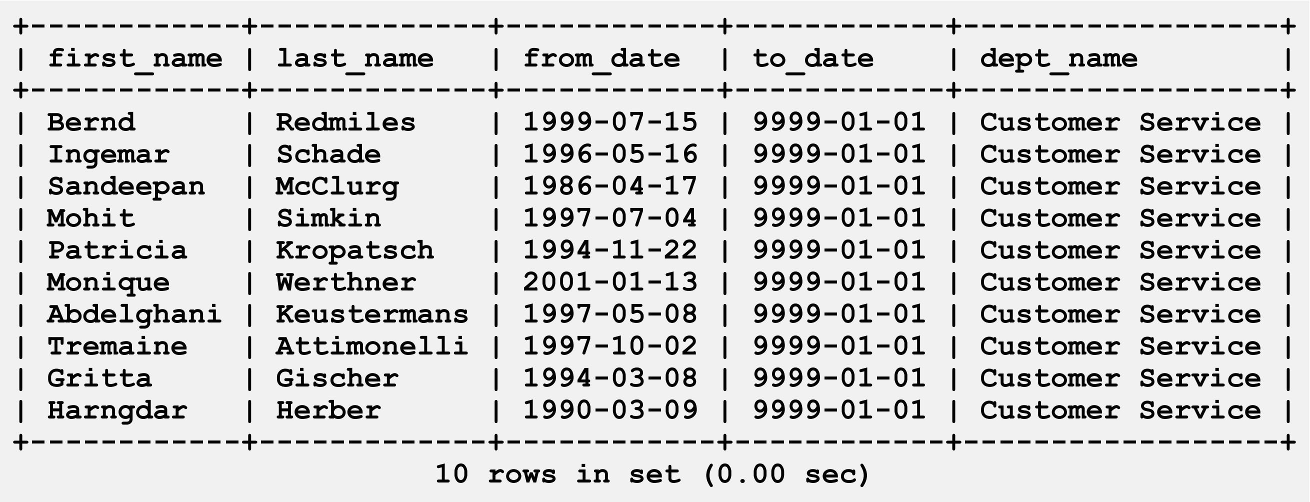
```
select a.first_name, a.last_name, b.from_date, b.to_date, c.dept_name
from employees as a, dept_emp as b, departments as c
where a.emp_no = b.emp_no and b.dept_no = c.dept_no
limit 10;
```

```
+------------+-------------+------------+------------+------------------+
| first_name | last_name   | from_date  | to_date    | dept_name        |
+------------+-------------+------------+------------+------------------+
| Bernd      | Redmiles    | 1999-07-15 | 9999-01-01 | Customer Service |
| Ingemar    | Schade      | 1996-05-16 | 9999-01-01 | Customer Service |
| Sandeepan  | McClurg     | 1986-04-17 | 9999-01-01 | Customer Service |
| Jiang      | Birjandi    | 1990-05-27 | 1996-09-21 | Customer Service |
| Mohit      | Simkin      | 1997-07-04 | 9999-01-01 | Customer Service |
| Patricia   | Kropatsch   | 1994-11-22 | 9999-01-01 | Customer Service |
| Monique    | Werthner    | 2001-01-13 | 9999-01-01 | Customer Service |
| Abdelghani | Keustermans | 1997-05-08 | 9999-01-01 | Customer Service |
| Tremaine   | Attimonelli | 1997-10-02 | 9999-01-01 | Customer Service |
| Gritta     | Gischer     | 1994-03-08 | 9999-01-01 | Customer Service |
+------------+-------------+------------+------------+------------------+
                        10 rows in set (0.00 sec)
```

Simple queries

Joining tables: employees and departments

```
select a.first_name, a.last_name, b.from_date, b.to_date, c.dept_name
from employees as a, dept_emp as b, departments as c
where a.emp_no = b.emp_no and b.dept_no = c.dept_no
   and b.from_date <= current_date and b.to_date >= current_date
limit 10;
```



```
+------------+-------------+------------+------------+------------------+
| first_name | last_name   | from_date  | to_date    | dept_name        |
+------------+-------------+------------+------------+------------------+
| Bernd      | Redmiles    | 1999-07-15 | 9999-01-01 | Customer Service |
| Ingemar    | Schade      | 1996-05-16 | 9999-01-01 | Customer Service |
| Sandeepan  | McClurg     | 1986-04-17 | 9999-01-01 | Customer Service |
| Mohit      | Simkin      | 1997-07-04 | 9999-01-01 | Customer Service |
| Patricia   | Kropatsch   | 1994-11-22 | 9999-01-01 | Customer Service |
| Monique    | Werthner    | 2001-01-13 | 9999-01-01 | Customer Service |
| Abdelghani | Keustermans | 1997-05-08 | 9999-01-01 | Customer Service |
| Tremaine   | Attimonelli | 1997-10-02 | 9999-01-01 | Customer Service |
| Gritta     | Gischer     | 1994-03-08 | 9999-01-01 | Customer Service |
| Harngdar   | Herber      | 1990-03-09 | 9999-01-01 | Customer Service |
+------------+-------------+------------+------------+------------------+
                      10 rows in set (0.00 sec)
```

Simple queries

Joining tables: employees and departments

```
select a.first_name, a.last_name, c.dept_name
from employees as a, dept_emp as b, departments as c
where a.emp_no = b.emp_no and b.dept_no = c.dept_no
    and b.from_date <= current_date and b.to_date >= current_date
limit 10;
```

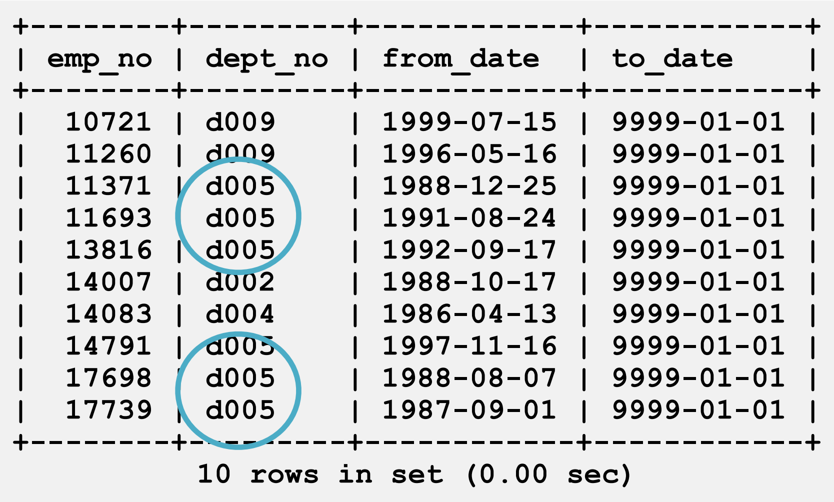
```
+------------+-------------+------------------+
| first_name | last_name   | dept_name        |
+------------+-------------+------------------+
| Bernd      | Redmiles    | Customer Service |
| Ingemar    | Schade      | Customer Service |
| Sandeepan  | McClurg     | Customer Service |
| Mohit      | Simkin      | Customer Service |
| Patricia   | Kropatsch   | Customer Service |
| Monique    | Werthner    | Customer Service |
| Abdelghani | Keustermans | Customer Service |
| Tremaine   | Attimonelli | Customer Service |
| Gritta     | Gischer     | Customer Service |
| Harngdar   | Herber      | Customer Service |
+------------+-------------+------------------+
            10 rows in set (0.00 sec)
```

Sorting

Sorting

Number of employees in each department

```
select *
from dept_emp
where from_date <= current_date
   and to_date >= current_date
limit 10;
```



```
+--------+---------+------------+------------+
| emp_no | dept_no | from_date  | to_date    |
+--------+---------+------------+------------+
|  10721 | d009    | 1999-07-15 | 9999-01-01 |
|  11260 | d009    | 1996-05-16 | 9999-01-01 |
|  11371 | d005    | 1988-12-25 | 9999-01-01 |
|  11693 | d005    | 1991-08-24 | 9999-01-01 |
|  13816 | d005    | 1992-09-17 | 9999-01-01 |
|  14007 | d002    | 1988-10-17 | 9999-01-01 |
|  14083 | d004    | 1986-04-13 | 9999-01-01 |
|  14791 | d005    | 1997-11-16 | 9999-01-01 |
|  17698 | d005    | 1988-08-07 | 9999-01-01 |
|  17739 | d005    | 1987-09-01 | 9999-01-01 |
+--------+---------+------------+------------+
           10 rows in set (0.00 sec)
```

Sorting

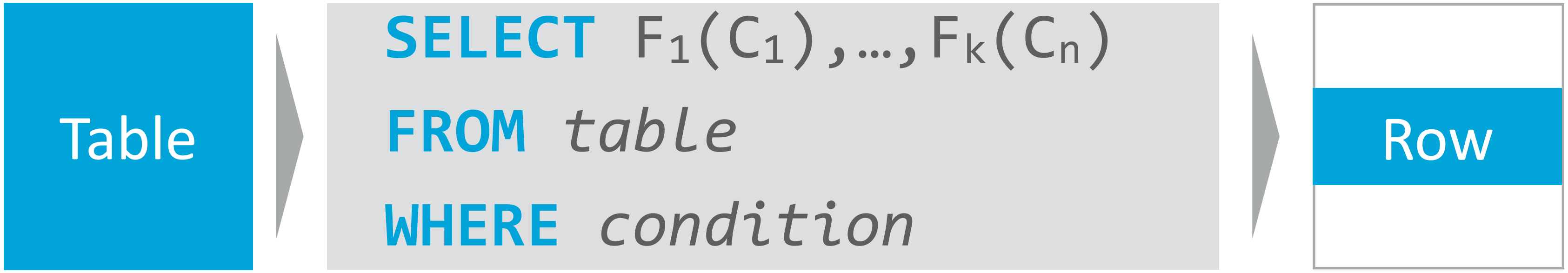
Number of employees in each department

```
select *
from dept_emp
where from_date <= current_date
    and to_date >= current_date
order by dept_no
limit 10;
```

```
+--------+---------+------------+------------+
| emp_no | dept_no | from_date  | to_date    |
+--------+---------+------------+------------+
|  21637 | d001    | 2000-01-14 | 9999-01-01 |
|  25949 | d001    | 1995-10-23 | 9999-01-01 |
|  44474 | d001    | 1996-09-11 | 9999-01-01 |
|  84372 | d001    | 1997-08-08 | 9999-01-01 |
| 102629 | d001    | 1989-07-21 | 9999-01-01 |
| 104349 | d001    | 1994-07-05 | 9999-01-01 |
| 110022 | d001    | 1985-01-01 | 9999-01-01 |
| 110039 | d001    | 1986-04-12 | 9999-01-01 |
| 205675 | d001    | 1990-11-16 | 9999-01-01 |
| 242872 | d001    | 1990-09-26 | 9999-01-01 |
+--------+---------+------------+------------+
10 rows in set (0.00 sec)
```

Introduction to Aggregation

Direct Aggregation Query Structure



SELECT $F_1(C_1),\ldots,F_k(C_n)$

FROM *table*

WHERE *condition*

Applies the functions to the values where the ***condition*** is true; returns only **one** row (with a column displaying the result of each aggregation function)

- $\mathbf{F_1, \ldots, F_k}$ are Aggregation Functions
- $\mathbf{C_1, \ldots, C_n}$ are columns of ***table***

Direct Aggregation

- *Find the number of customers in the bank*

```
SELECT COUNT(*)
FROM customer;
```

```
 count
-------
     15
```

- *Find the number of depositors in the bank*

```
SELECT COUNT(DISTINCT customer_name)
FROM depositor;
```

```
 count
-------
      7
```

- *Find the average account balance at the 'Central' branch*

```
SELECT AVG(balance)
FROM account
WHERE branch_name = 'Central'
```

```
    avg
----------
 650.00
```

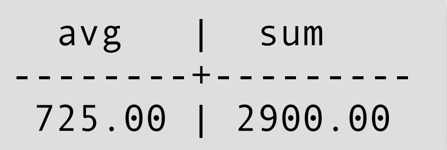
Aggregate Functions

- `COUNT([DISTINCT] A)`
  - The count of (distinct) values on column A
- `SUM([DISTINCT] A)`
  - The sum of (distinct) values on column A
- `AVG([DISTINCT] A)`
  - The average of (distinct) values on column A
- `MAX(A)`
  - The maximum value on column A
- `MIN(A)`
  - The minimum value on column A

Aggregate Functions

Find the average balance of accounts, and the sum of balances at the branches ‘Central’ or ‘Uptown’

```
SELECT AVG(balance), SUM(balance)
FROM account
WHERE branch_name = 'Central'
   OR branch_name = 'Uptown'
```



```
  avg    |   sum
---------+----------
 725.00 | 2900.00
```

Only one row!

Aggregate Functions

- Find the number of persons

```
SELECT COUNT(*)
FROM person;
```

```
 count
-------
 50000
```

- Find the number of female persons

```
SELECT COUNT(*)
FROM person
WHERE gender = 'female';
```

```
 count
-------
 24803
```

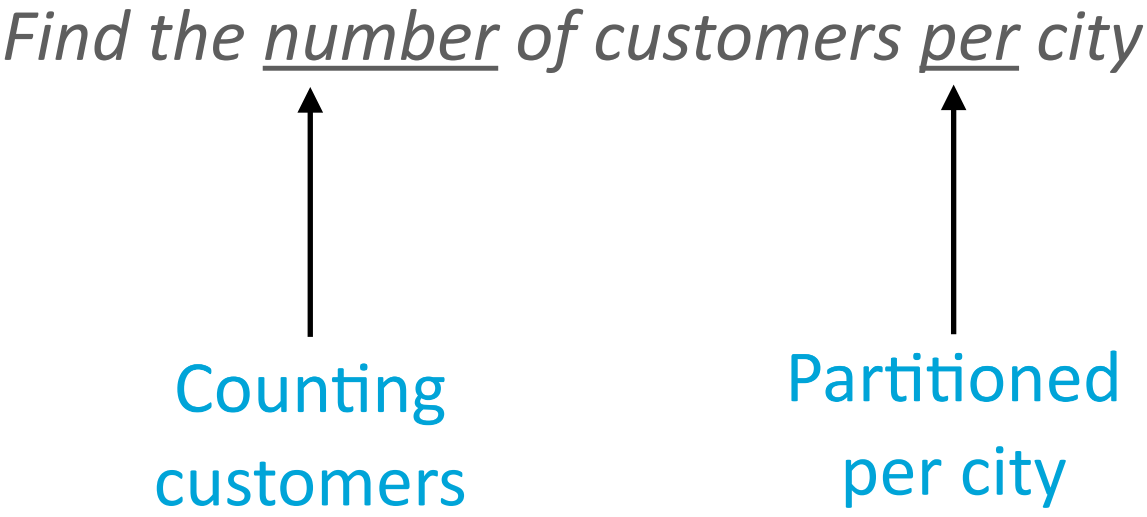
- Find the average weight of persons that live in Portugal

```
SELECT AVG(kilograms)
FROM person
WHERE country = 'Portugal'
```

```
    avg
----------
  80.056
```

Introduction to Grouping and Aggregation

Partitioned Aggregation



Find the number of customers per city

Counting customers

Partitioned per city

```
SELECT COUNT(*)
FROM customer
```

```
GROUP BY customer_city
```

Finding Partitions

```
SELECT *  FROM customer
```

```
 customer_name | customer_street | customer_city
---------------+-----------------+----------------
 Adams         | Main Street     | Lisbon
 Brown         | Main
 Cook          | Main
 Davis         | Church Street   | Oporto
 Evans         | Fore
 Flores        | Stat
 Gonzalez      | Sunn
 Iacocca       | Spri
 Johnson       | New
 King          | Gard
 Lopez         | Gran
 Martin        | Roya
 Nguyen        | Scho
 Oliver        | Firs
 Parker        | Hope
```

```
SELECT *  FROM customer ORDER BY customer_city
```

```
 customer_name | customer_street | customer_city
---------------+-----------------+----------------
 King          | Garden Street   | Aveiro
 Flores        | Station Street  | Braga
 Martin        | Royal Street    | Braga
 Johnson       | New Street      | Cascais
 Nguyen        | School Street   | Castelo Branco
 Iacocca       | Spring Steet    | Coimbra
 Evans         | Forest Street   | Coimbra
 Gonzalez      | Sunny Street    | Faro
 Adams         | Main Street     | Lisbon
 Cook          | Main Street     | Lisbon
 Davis         | Church Street   | Oporto
 Brown         | Main Street     | Oporto
 Oliver        | First Stret     | Oporto
 Parker        | Hope Street     | Oporto
 Lopez         | Grand Street    | Vila Real
```

Partitioned Aggregation: Step-by-step

```
SELECT *
FROM customer
ORDER BY customer_city
```

```
 customer_name | customer_street | customer_city
---------------+-----------------+----------------
 King          | Garden Street   | Aveiro
 Flores        | Station Street  | Braga
 Martin        | Royal Street    | Braga
 Johnson       | New Street      | Cascais
 Nguyen        | School Street   | Castelo Branco
 Iacocca       | Spring Steet    | Coimbra
 Evans         | Forest Street   | Coimbra
 Gonzalez      | Sunny Street    | Faro
 Adams         | Main Street     | Lisbon
 Cook          | Main Street     | Lisbon
 Davis         | Church Street   | Oporto
 Brown         | Main Street     | Oporto
 Oliver        | First Stret     | Oporto
 Parker        | Hope Street     | Oporto
 Lopez         | Grand Street    | Vila Real
```

```
SELECT COUNT(*)
FROM customer
```

```
 count
-------
    15
```

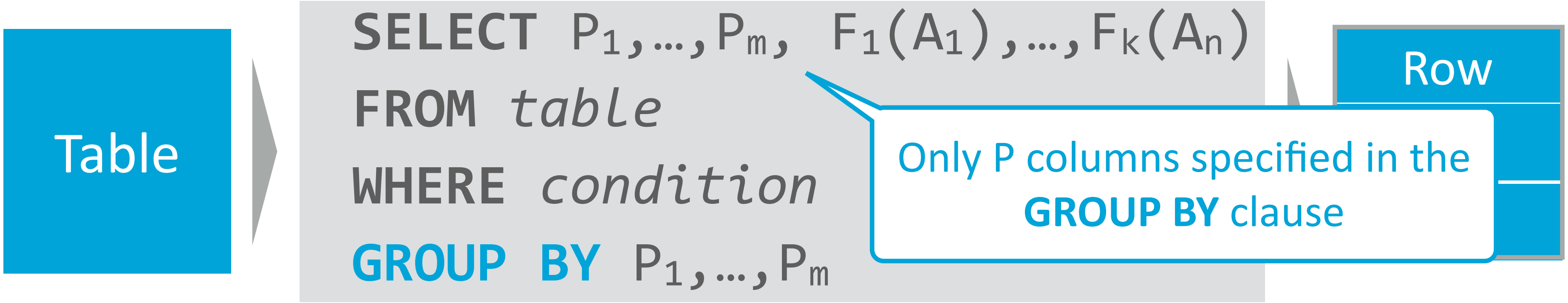
```
SELECT COUNT(*)
FROM customer
GROUP BY customer_city
```

```
  count
---------
    1
    2
    1
    1
    2
    1
    2
    4
    1
```

```
SELECT customer_city, COUNT(*)
FROM customer
GROUP BY customer_city
```

```
customer_city | count
--------------+------
Aveiro        |     1
Braga         |     2
Cascais       |     1
Castelo Branco|     1
Coimbra       |     2
Faro          |     1
Lisbon        |     2
Oporto        |     4
Vila Real     |     1
```

Partitioned Aggregation Query Structure



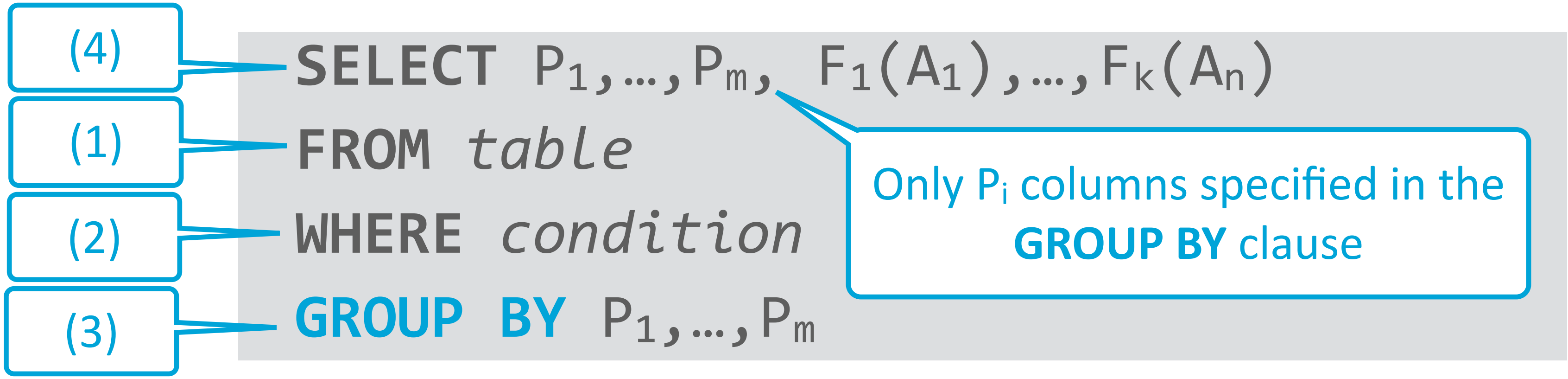
```
SELECT P1,…,Pm, F1(A1),…,Fk(An)
FROM table
WHERE condition
GROUP BY P1,…,Pm
```

Only P columns specified in the **GROUP BY** clause

Applies the functions to partitions (sub-tables) that agree on the combinations of values of $P_1,…,P_m$; returns multiple rows (one for each partition)

- **F_1, …, F_k** are Aggregation Functions
- **A_1, …, A_n** are the aggregated columns (of ***table***)
- **P_1, …, P_m** are partitioning columns (of ***table***)

The **GROUP BY** clause



(4) **SELECT** $P_1,…,P_m,$ $F_1(A_1),…,F_k(A_n)$

(1) **FROM** *table*

(2) **WHERE** *condition*

(3) **GROUP BY** $P_1,…,P_m$

Only P_i columns specified in the **GROUP BY** clause

NOTE: The combinations of values of P_1, ..., P_m are expected to have duplicates in the *table*

Selecting an attribute that is not partitioned

Find the name of customers per city

↑↑↑ This is not an aggregate query (why?)🤔

```
SELECT customer_name, COUNT(*)
FROM customer
GROUP BY customer_city;
```

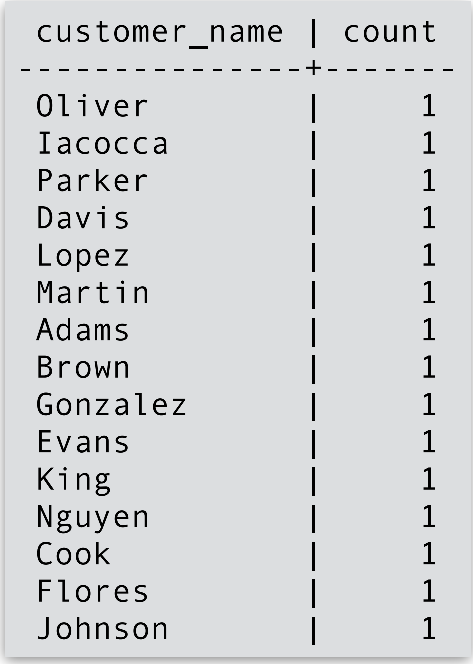
↑↑↑ Projects an attribute that does not exist

Selecting an attribute that is not partitioned is an error!🙀🙀

Partitioning by an column with unique values

What happens when the partitioning is made on a filed that only has distinct values?

```
SELECT customer_name, COUNT(*)
FROM customer
GROUP BY customer_name;
```



```
 customer_name | count
---------------+-------
 Oliver        |     1
 Iacocca       |     1
 Parker        |     1
 Davis         |     1
 Lopez         |     1
 Martin        |     1
 Adams         |     1
 Brown         |     1
 Gonzalez      |     1
 Evans         |     1
 King          |     1
 Nguyen        |     1
 Cook          |     1
 Flores        |     1
 Johnson       |     1
```

Each group (partition) will have only one row!

⚠ We are grouping by an attribute (or attibutes) that are a candidate key

Aggregate Filtering

Find the names of all branches where the average account balance is above 750€

```
SELECT branch_name, AVG(balance)
FROM account
WHERE AVG(balance) > 750;
GROUP BY branch_name
```

```
SELECT *
FROM (
     SELECT branch_name, AVG(balance)
     FROM account
     GROUP BY branch_name
)
WHERE balance > 750;
```

Aggregate Filtering

Find the names of all branches where the average account balance is above than 750€

```
SELECT branch_name, AVG(balance)
FROM account
GROUP BY branch_name
HAVING AVG(balance) > 750;
```

```
 branch_name |   avg
-------------+-----------
 Uptown      | 800.00
 Round Hill  | 800.00
```

NOTE: conditions in the **HAVING** clause are applied *after* forming groups while conditions in the **WHERE** clause are applied *before* forming groups!

Example: Aggregate Filtering

What are the branches with at least two accounts?

```
SELECT branch_name, COUNT(*)
FROM account
GROUP BY branch_name
HAVING COUNT(*) >= 2;
```

```
 branch_name | count
-------------+-------
 Central     |     2
 Uptown      |     2
 Downtown    |     2
```

```
SELECT branch_name
FROM account
GROUP BY branch_name
HAVING COUNT(*) >= 2;
```

```
 branch_name
--------------------
 Central
 Uptown
 Downtown
```

The **HAVING** clause

A SELECT statement with a HAVING clause has the form:

(5) **SELECT** $P_1,\dots,P_m$, $F_1(A_1),\dots,F_k(A_n)$
(1) **FROM** *table*
(2) **WHERE** *condition*
(3) **GROUP BY** $P_1,\dots,P_m$
(4) **HAVING** *aggregate_condition*

- Where ***aggregate_condition*** is a condition with aggregate functions

Grouping and Aggregation

Grouping and aggregation

Number of employees in each department

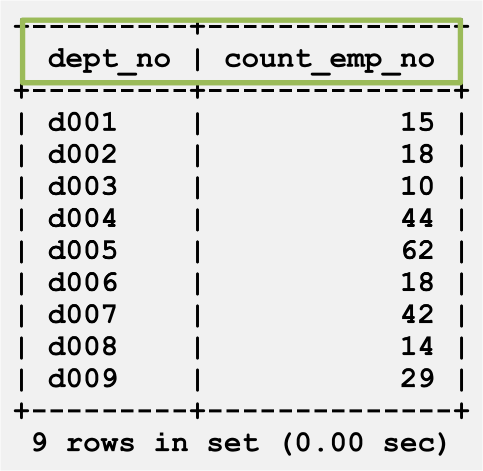
```
select dept_no, count(emp_no)
from dept_emp
where from_date <= current_date
    and to_date >= current_date
group by dept_no;
```

```
+---------+---------------+
| dept_no | count(emp_no) |
+---------+---------------+
| d001    |            15 |
| d002    |            18 |
| d003    |            10 |
| d004    |            44 |
| d005    |            62 |
| d006    |            18 |
| d007    |            42 |
| d008    |            14 |
| d009    |            29 |
+---------+---------------+
 9 rows in set (0.01 sec)
```

Grouping and aggregation

Number of employees in each department

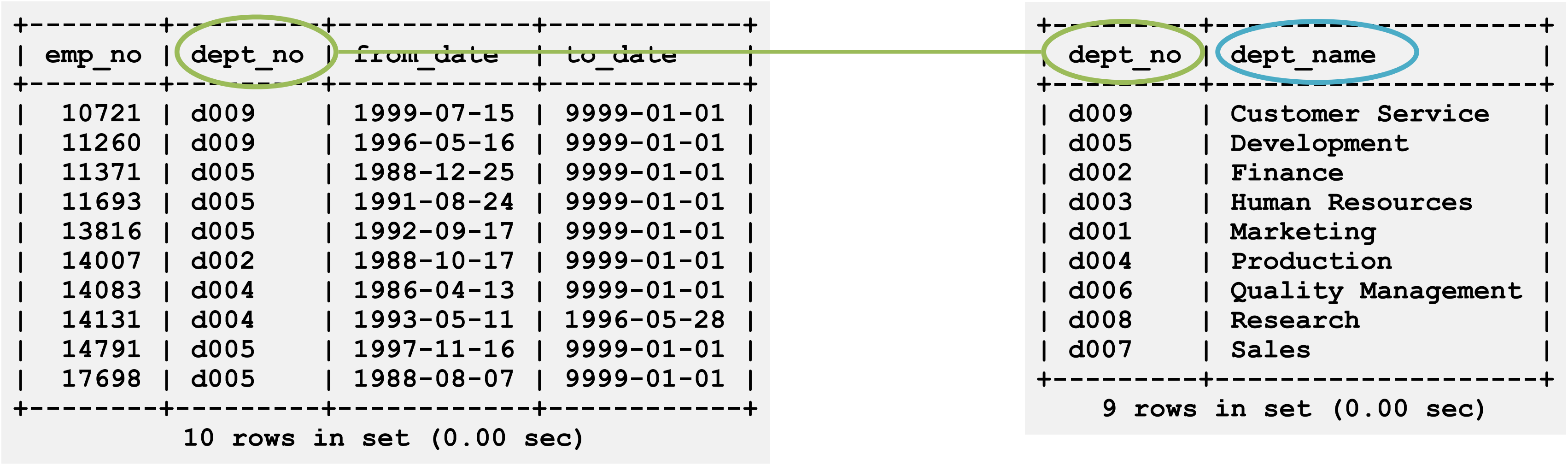
```
select dept_no, count(emp_no) as count_emp_no
from dept_emp
where from_date <= current_date
    and to_date >= current_date
group by dept_no;
```



```
+---------+--------------+
| dept_no | count_emp_no |
+---------+--------------+
| d001    |           15 |
| d002    |           18 |
| d003    |           10 |
| d004    |           44 |
| d005    |           62 |
| d006    |           18 |
| d007    |           42 |
| d008    |           14 |
| d009    |           29 |
+---------+--------------+
9 rows in set (0.00 sec)
```

Grouping and aggregation

Number of employees in each department



dept_emp

```
+--------+---------+------------+------------+
| emp_no | dept_no | from_date  | to_date    |
+--------+---------+------------+------------+
|  10721 | d009    | 1999-07-15 | 9999-01-01 |
|  11260 | d009    | 1996-05-16 | 9999-01-01 |
|  11371 | d005    | 1988-12-25 | 9999-01-01 |
|  11693 | d005    | 1991-08-24 | 9999-01-01 |
|  13816 | d005    | 1992-09-17 | 9999-01-01 |
|  14007 | d002    | 1988-10-17 | 9999-01-01 |
|  14083 | d004    | 1986-04-13 | 9999-01-01 |
|  14131 | d004    | 1993-05-11 | 1996-05-28 |
|  14791 | d005    | 1997-11-16 | 9999-01-01 |
|  17698 | d005    | 1988-08-07 | 9999-01-01 |
+--------+---------+------------+------------+
           10 rows in set (0.00 sec)
```

departments

```
+---------+--------------------+
| dept_no | dept_name          |
+---------+--------------------+
| d009    | Customer Service   |
| d005    | Development        |
| d002    | Finance            |
| d003    | Human Resources    |
| d001    | Marketing          |
| d004    | Production         |
| d006    | Quality Management |
| d008    | Research           |
| d007    | Sales              |
+---------+--------------------+
    9 rows in set (0.00 sec)
```

Grouping and aggregation

Number of employees in each department

```
select *
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
limit 10;
```

```
+--------+---------+------------+------------+---------+------------------+
| emp_no | dept_no | from_date  | to_date    | dept_no | dept_name        |
+--------+---------+------------+------------+---------+------------------+
|  10721 | d009    | 1999-07-15 | 9999-01-01 | d009    | Customer Service |
|  11260 | d009    | 1996-05-16 | 9999-01-01 | d009    | Customer Service |
|  22772 | d009    | 1986-04-17 | 9999-01-01 | d009    | Customer Service |
|  30297 | d009    | 1990-05-27 | 1996-09-21 | d009    | Customer Service |
|  40234 | d009    | 1997-07-04 | 9999-01-01 | d009    | Customer Service |
|  61712 | d009    | 1994-11-22 | 9999-01-01 | d009    | Customer Service |
|  64387 | d009    | 2001-01-13 | 9999-01-01 | d009    | Customer Service |
|  68491 | d009    | 1997-05-08 | 9999-01-01 | d009    | Customer Service |
|  73133 | d009    | 1997-10-02 | 9999-01-01 | d009    | Customer Service |
|  91524 | d009    | 1994-03-08 | 9999-01-01 | d009    | Customer Service |
+--------+---------+------------+------------+---------+------------------+
                        10 rows in set (0.00 sec)
```

Grouping and aggregation

Number of employees in each department

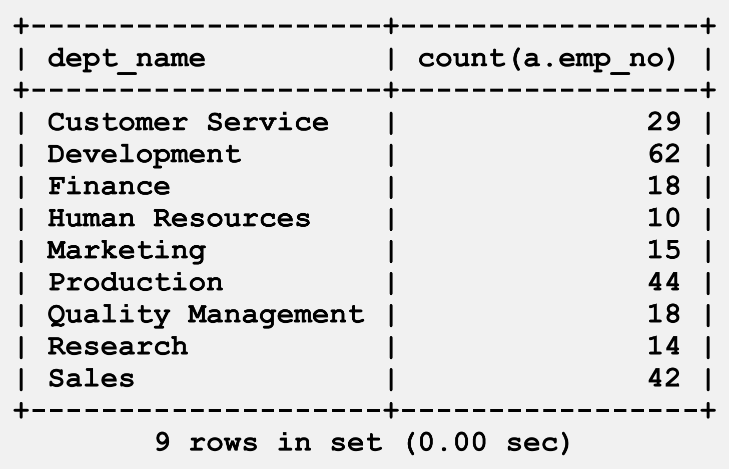
```
select *
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
limit 10;
```

```
+--------+---------+------------+------------+---------+------------------+
| emp_no | dept_no | from_date  | to_date    | dept_no | dept_name        |
+--------+---------+------------+------------+---------+------------------+
|  10721 | d009    | 1999-07-15 | 9999-01-01 | d009    | Customer Service |
|  11260 | d009    | 1996-05-16 | 9999-01-01 | d009    | Customer Service |
|  22772 | d009    | 1986-04-17 | 9999-01-01 | d009    | Customer Service |
|  40234 | d009    | 1997-07-04 | 9999-01-01 | d009    | Customer Service |
|  61712 | d009    | 1994-11-22 | 9999-01-01 | d009    | Customer Service |
|  64387 | d009    | 2001-01-13 | 9999-01-01 | d009    | Customer Service |
|  68491 | d009    | 1997-05-08 | 9999-01-01 | d009    | Customer Service |
|  73133 | d009    | 1997-10-02 | 9999-01-01 | d009    | Customer Service |
|  91524 | d009    | 1994-03-08 | 9999-01-01 | d009    | Customer Service |
| 103361 | d009    | 1990-03-09 | 9999-01-01 | d009    | Customer Service |
+--------+---------+------------+------------+---------+------------------+
                    10 rows in set (0.00 sec)
```

Grouping and aggregation

Number of employees in each department

```
select b.dept_name, count(a.emp_no)
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by b.dept_name;
```



```
+--------------------+-----------------+
| dept_name          | count(a.emp_no) |
+--------------------+-----------------+
| Customer Service   |              29 |
| Development        |              62 |
| Finance            |              18 |
| Human Resources    |              10 |
| Marketing          |              15 |
| Production         |              44 |
| Quality Management |              18 |
| Research           |              14 |
| Sales              |              42 |
+--------------------+-----------------+
      9 rows in set (0.00 sec)
```

Grouping and aggregation

Number of employees in each department (alternative)

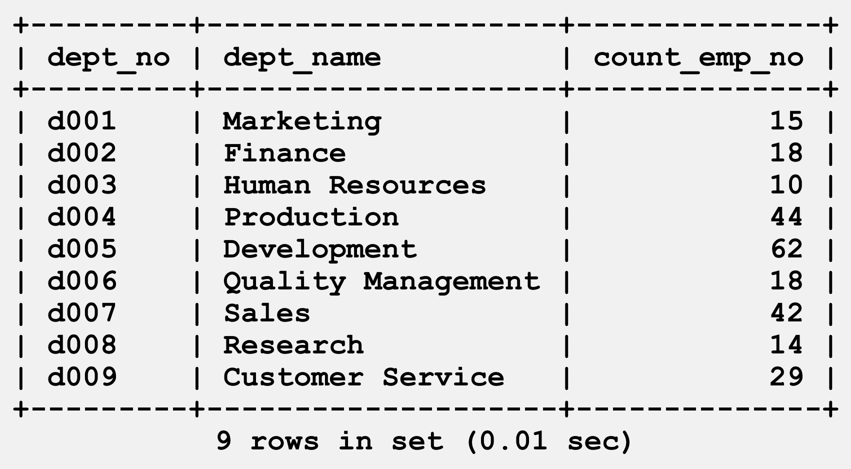
```
select a.dept_no, b.dept_name, count(a.emp_no)
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name;
```

```
+---------+--------------------+-----------------+
| dept_no | dept_name          | count(a.emp_no) |
+---------+--------------------+-----------------+
| d001    | Marketing          |              15 |
| d002    | Finance            |              18 |
| d003    | Human Resources    |              10 |
| d004    | Production         |              44 |
| d005    | Development        |              62 |
| d006    | Quality Management |              18 |
| d007    | Sales              |              42 |
| d008    | Research           |              14 |
| d009    | Customer Service   |              29 |
+---------+--------------------+-----------------+
             9 rows in set (0.00 sec)
```

Grouping and aggregation

Number of employees in each department

```
select a.dept_no, b.dept_name, count(a.emp_no) as count_emp_no
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name;
```



```
+---------+--------------------+--------------+
| dept_no | dept_name          | count_emp_no |
+---------+--------------------+--------------+
| d001    | Marketing          |           15 |
| d002    | Finance            |           18 |
| d003    | Human Resources    |           10 |
| d004    | Production         |           44 |
| d005    | Development        |           62 |
| d006    | Quality Management |           18 |
| d007    | Sales              |           42 |
| d008    | Research           |           14 |
| d009    | Customer Service   |           29 |
+---------+--------------------+--------------+
            9 rows in set (0.01 sec)
```

Grouping and sorting

Number of employees in each department

```
select a.dept_no, b.dept_name, count(a.emp_no) as count_emp_no
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name
order by b.dept_name;
```

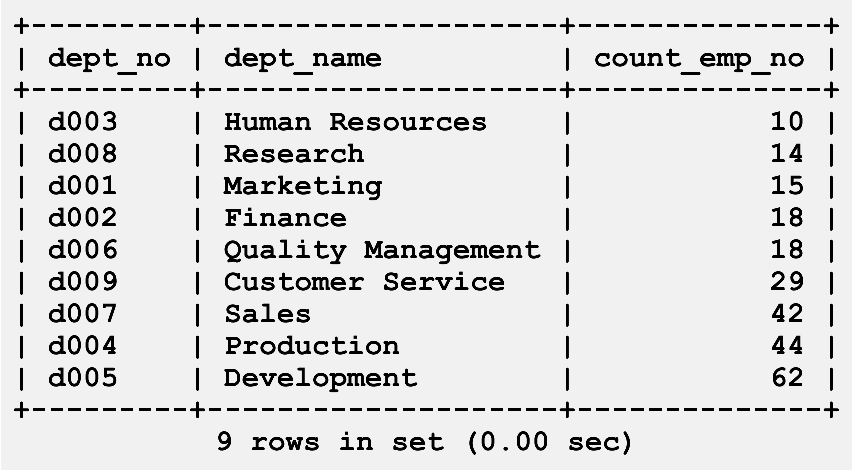
```
+---------+--------------------+--------------+
| dept_no | dept_name          | count_emp_no |
+---------+--------------------+--------------+
| d009    | Customer Service   |           29 |
| d005    | Development        |           62 |
| d002    | Finance            |           18 |
| d003    | Human Resources    |           10 |
| d001    | Marketing          |           15 |
| d004    | Production         |           44 |
| d006    | Quality Management |           18 |
| d008    | Research           |           14 |
| d007    | Sales              |           42 |
+---------+--------------------+--------------+
            9 rows in set (0.00 sec)
```

Grouping and Aggregation, and Filtering

Grouping, aggregation and sorting

Number of employees in each department

```
select a.dept_no, b.dept_name, count(a.emp_no) as count_emp_no
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name
order by count_emp_no;
```



```
+---------+--------------------+--------------+
| dept_no | dept_name          | count_emp_no |
+---------+--------------------+--------------+
| d003    | Human Resources    |           10 |
| d008    | Research           |           14 |
| d001    | Marketing          |           15 |
| d002    | Finance            |           18 |
| d006    | Quality Management |           18 |
| d009    | Customer Service   |           29 |
| d007    | Sales              |           42 |
| d004    | Production         |           44 |
| d005    | Development        |           62 |
+---------+--------------------+--------------+
           9 rows in set (0.00 sec)
```

Grouping, aggregation and sorting

Number of employees in each department

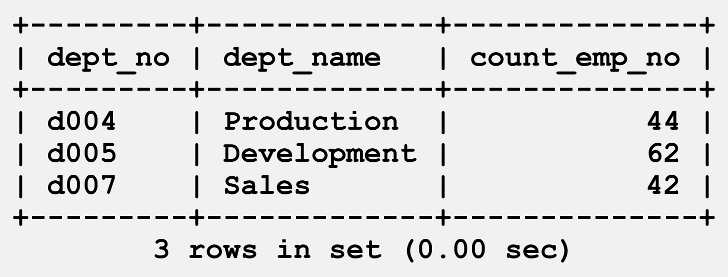
```
select a.dept_no, b.dept_name, count(a.emp_no) as count_emp_no
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name
order by count_emp_no desc;
```

```
+---------+--------------------+--------------+
| dept_no | dept_name          | count_emp_no |
+---------+--------------------+--------------+
| d005    | Development        |           62 |
| d004    | Production         |           44 |
| d007    | Sales              |           42 |
| d009    | Customer Service   |           29 |
| d002    | Finance            |           18 |
| d006    | Quality Management |           18 |
| d001    | Marketing          |           15 |
| d008    | Research           |           14 |
| d003    | Human Resources    |           10 |
+---------+--------------------+--------------+
9 rows in set (0.00 sec)
```

Grouping, aggregation and filtering

Number of employees in each department

```
select a.dept_no, b.dept_name, count(a.emp_no) as count_emp_no
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name
having count_emp_no >= 40;
```



```
+---------+-------------+--------------+
| dept_no | dept_name   | count_emp_no |
+---------+-------------+--------------+
| d004    | Production  |           44 |
| d005    | Development |           62 |
| d007    | Sales       |           42 |
+---------+-------------+--------------+
        3 rows in set (0.00 sec)
```

Grouping, aggregation, filtering and sorting

Number of employees in each department

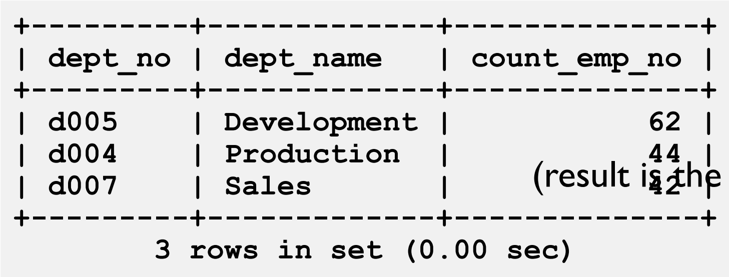
```
select a.dept_no, b.dept_name, count(a.emp_no) as count_emp_no
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name
having count_emp_no >= 40
order by b.dept_name;
```

```
+---------+-------------+--------------+
| dept_no | dept_name   | count_emp_no |
+---------+-------------+--------------+
| d005    | Development |           62 |
| d004    | Production  |           44 |
| d007    | Sales       |           42 |
+---------+-------------+--------------+
        3 rows in set (0.00 sec)
```

Grouping, aggregation, filtering and sorting

Number of employees in each department

```
select a.dept_no, b.dept_name, count(a.emp_no) as count_emp_no
from dept_emp as a, departments as b
where a.dept_no = b.dept_no
    and a.from_date <= current_date
    and a.to_date >= current_date
group by a.dept_no, b.dept_name
having count_emp_no >= 40
order by count_emp_no desc;
```



```
+---------+-------------+--------------+
| dept_no | dept_name   | count_emp_no |
+---------+-------------+--------------+
| d005    | Development |           62 |
| d004    | Production  |           44 |
| d007    | Sales       |           42 |
+---------+-------------+--------------+
        3 rows in set (0.00 sec)
```

(result is the same, by coincidence)

TÉCNICO LISBOA