



Очистка данных

Очистка данных играет важную роль в процессе загрузки данных. В задачи дата инженера входит привести данные к одному виду, оставить только валидные данные и отбросить все лишнее.

Для очистки данных в SQL есть множество строковых функций.

Строковые функции (1 набор)

<u>CONCAT</u>	 CONCAT('A','BC')	 объединение строк
<u>INSTR</u>	INSTR('This is a playlist', 'is')	возвращает позицию, где начинается подстрока
<u>LENGTH</u>	LENGTH('ABC')	возвращает длину строки
<u>REPLACE</u>	REPLACE('JACK AND JOND','J','BL')	заменяет подстроку на другую
<u>SUBSTR</u>	SUBSTR('Oracle Substring', 1, 6)	позволяет получить подстроку

Примеры

```
select concat('привет', 'друг') from dual; -- конкатенация
select 'привет' || ' ' || 'друг' from dual; -- конкатенация через оператор ||

-- позиция подстроки в строке
select instr('привет мой друг', 'мой') from dual;
-- при отсутствии подстроки значение 0
select instr('привет мой друг', 'мой_') from dual;

-- длина строки
select length('привет мой друг') from dual;
-- при измерении длины учитываются боковые пробельные символы
select length('привет мой друг          ') from dual;

-- замена подстроки на другую подстроку
select replace('привет мой друг', 'друг', 'товарищ') from dual;
select replace('привет мой друг', ' ', '_') from dual;

-- достать подстроку из строки
-- 1 параметр - позиция с которой начинается подстрока
```

```
-- 2 параметр - длина получаемой подстроки
select substr('привет мой друг', 8, 3) from dual;
```

Задание

Таблица **hr.employees**

1. Сформируйте поле, которое указывает последние 4 цифры номера телефона сотрудника
2. Уберите в телефоне символ .
3. Определите работников с самым длинным email адресом
4. Сформируйте поле с именем и фамилией сотрудника

```
-- Сформируйте поле, которое указывает последние 4 цифры номера телефона сотрудника

select substr(PHONE_NUMBER,length(PHONE_NUMBER)-3 ) from hr.employees;

-- Уберите в телефоне символ .

select replace(PHONE_NUMBER, '.', '') from hr.employees;
-- в случае если подстрока меняется на пустую строку, второй параметр не обязателен
select replace(PHONE_NUMBER, '.') from hr.employees;

-- Определите работников с самым длинным email адресом

select * from hr.employees
where length(email) = (
    select max(length(email)) from hr.employees;
);

-- Сформируйте поле с именем и фамилией сотрудника
select FIRST_NAME || ' ' || LAST_NAME from hr.employees;
```

Строковые функции (2 набор)

<u>Aa</u> LOWER	⇄ LOWER('Abc')	⇄ Перевод в верхний регистр
<u>UPPER</u>	UPPER('Abc')	Перевод в нижний регистр

<u>Aa</u> LOWER	≡ LOWER('Abc')	≡ Перевод в верхний регистр
<u>INITCAP</u>	INITCAP('hi there')	Первая буква большая, остальные маленькие

Примеры

```
select
  lower('пРиВеТ'),
  upper('пРиВеТ'),
  initcap('пРиВеТ')
from dual;
```

Строковые функции (3 набор)

<u>Aa</u> LPAD	≡ LPAD('ABC',5,'*')	≡ Добавляет указанное кол-во символов слева
<u>RPAD</u>	RPAD('ABC',5,'*')	Добавляет указанное кол-во символов справа
<u>LTRIM</u>	LTRIM(' ABC ')	Убирает пробелы слева
<u>RTRIM</u>	LTRIM(' ABC ')	Убирает пробелы справа
<u>TRIM</u>	LTRIM(' ABC ')	Убирает пробелы по краям

Примеры

```
select
  lpad(n, 10, '0'),
  rpad(n, 10, '0')
from (
  select 12 as n from dual
  union all
  select 1232 as n from dual
  union all
  select 6533 as n from dual
  union all
  select 22145 as n from dual
) t1;

select
  ltrim(n),
  rtrim(n),
  trim(n) from (
  select
    '      привет друг      ' as n
  from dual
) t1
```

