

SQL实例解析

软件学院



Part 1 字符串处理

- 1.1 遍历字符串
- 1.2 嵌入引号
- 1.3 统计字符出现的次数
- 1.4 删除不想要的字符
- 1.5 分离数字和字符数据
- 1.6 判断含有字母和数字的字符串
- 1.7 提取姓名的首字母



SQL的字符串处理

- SQL并不专门用于处理复杂的字符串
 - 很多时候非常麻烦，令人沮丧
 - BUT，仍然有很多很好用的内置函数
- 任何事物，包括SQL都有自己好的一面和坏的令人厌恶的一面



1.1 遍历字符串

- 这是一切字符串处理的基础，你需要有逐字遍历字符串的能力
- SQL没有Loop循环功能，我们需要有**数据透视表** (T1, T10, T100...)
- **问题：**把EMP表中的ENAME=KING的字符串拆开显示为4行，每行一个字符

C
-
K
I
N
G

```
1 select substr(e.ename,iter.pos,1) as C
2   from (select ename from emp where ename = 'KING') e,
3        (select id as pos from t10) iter
4  where iter.pos <= length(e.ename)
```



1.1 遍历字符串（讨论）

完整的笛卡尔积

```
select ename, iter.pos
from (select ename from emp where ename = 'KING') e,
     (select id as pos from t10) iter
```

ENAME	POS
KING	1
KING	2
KING	3
KING	4
KING	5
KING	6
KING	7
KING	8
KING	9
KING	10

返回部分数据的笛卡尔积

```
select ename, iter.pos
from (select ename from emp where ename = 'KING') e,
     (select id as pos from t10) iter
where iter.pos <= length(e.ename)
```

ENAME	POS
KING	1
KING	2
KING	3
KING	4

把T10表作为数据透视表是常见的技巧



1.1 遍历字符串 (讨论)

想显示如下结果怎么办?

A	B
KING	G
ING	NG
NG	ING
G	KING

```
select substr(e.ename,iter.pos) a,  
       substr(e.ename,length(e.ename)-iter.pos+1) b  
from (select ename from emp where ename = 'KING') e,  
     (select id pos from t10) iter  
where iter.pos <= length(e.ename)
```

遍历字符串，是SQL字符串处理一切的起点



1.2 嵌入引号

- 如果想在字符串常量中嵌入引号，并且希望使用SQL产生如下所示的结果

QMARKS

g'day mate
beavers' teeth
,

```
1 select 'g'day mate' qmarks from t1 union all
2 select 'beavers" teeth'  from t1 union all
3 select '""               from t1
```



1.3 统计字符出现的次数

- **问题：**统计字符串中有多少个逗号？ 10,CLARK,MANAGER

```
1 select (length('10,CLARK,MANAGER')-  
2    length(replace('10,CLARK,MANAGER',','))) / length(',')  
3    as cnt  
4 from t1
```



1.3 统计字符出现的次数（讨论）

- **思考：**如何统计HELLO HELLO中出现了多少个LL

```
select
    (length('HELLO HELLO')-
    length(replace('HELLO HELLO','LL','')))/length('LL')
    as correct_cnt,
    (length('HELLO HELLO')-
    length(replace('HELLO HELLO','LL',''))) as incorrect_cnt
from t1
```

CORRECT_CNT	INCORRECT_CNT
2	4

除法运算是得到“次数”这类运算正确答案必须使用的运算手段



1.4 删除不想要的字符

- **问题：**从数据里删除指定的字符，从左边的结果集中的数据里删除所有的0和元音字母，并将删除后的值显示在STRIPPED1列和STRIPPED2列中，形成右边的结果集形态

ENAME	SAL
-----	-----
SMITH	800
ALLEN	1600
WARD	1250
JONES	2975
MARTIN	1250
BLAKE	2850
CLARK	2450
SCOTT	3000
KING	5000
TURNER	1500
ADAMS	1100
JAMES	950
FORD	3000
MILLER	1300

ENAME	STRIPPED1	SAL	STRIPPED2
-----	-----	-----	-----
SMITH	SMTH	800	8
ALLEN	LLN	1600	16
WARD	WRD	1250	125
JONES	JNS	2975	2975
MARTIN	MRTN	1250	125
BLAKE	BLK	2850	285
CLARK	CLRK	2450	245
SCOTT	SCTT	3000	3
KING	KNG	5000	5
TURNER	TRNR	1500	15
ADAMS	DMS	1100	11
JAMES	JMS	950	95
FORD	FRD	3000	3
MILLER	MLLR	1300	13



1.4 删除不想要的字符

Oracle and PostgreSQL

```
1 select ename,  
2    replace(translate(ename,'AEIOU','aaaaa'),'a')  
3    as stripped1,  
4    sal,  
5    replace(sal,0,'') as stripped2  
6 from emp
```

先用TRANSLATE函数把元音字母替换成一个特殊的字符,
然后使用REPLACE函数删除这个特殊字符

没有TRANSLATE函数, 那就只有做苦力活了

MySQL and SQL Server

```
1 select ename,  
2    replace(  
3    replace(  
4    replace(  
5    replace(  
6    replace(ename,'A',''),'E',''),'I',''),'O',''),'U','')  
7    as stripped1,  
8    sal,  
9    replace(sal,0,'') stripped2  
10 from emp
```



1.5 分离数字和字符数据

- **问题：** 把数据中的数字数据和字符数据分开，怎么办？

DATA	ENAME	SAL
-----	-----	-----
SMITH800	SMITH	800
ALLEN1600	ALLEN	1600
WARD1250	WARD	1250
JONES2975	JONES	2975
MARTIN1250	MARTIN	1250
BLAKE2850	BLAKE	2850
CLARK2450	CLARK	2450
SCOTT3000	SCOTT	3000
KING5000	KING	5000
TURNER1500	TURNER	1500
ADAMS1100	ADAMS	1100
JAMES950	JAMES	950
FORD3000	FORD	3000
MILLER1300	MILLER	1300



1.5 分离数字和字符数据

Oracle

```
1 select replace(  
2   translate(data,'0123456789','0000000000'),'0') ename,  
3   to_number(  
4     replace(  
5       translate(lower(data),  
6         'abcdefghijklmnopqrstuvwxyz',  
7         rpad('z',26,'z')),'z')) sal  
9   from (  
10  select ename||sal data  
11  from emp  
12  )
```

MySQL只用REPLACE怎么办?



1.6 判断含有字母和数字的字符

- **问题：**从表里筛选出部分行数据，筛选条件是只包含字母和数字字符

视图V

DATA

CLARK

KING

MILLER

SMITH, \$800.00

JONES, \$2975.00

SCOTT, \$3000.00

ADAMS, \$1100.00

FORD, \$3000.00

ALLEN30

WARD30

MARTIN30

BLAKE30

TURNER30

JAMES30



DATA

CLARK

KING

MILLER

ALLEN30

WARD30

MARTIN30

BLAKE30

TURNER30

JAMES30



1.6 判断含有字母和数字的字符

MySQL

```
1 select data
2 from V
3 where data regexp '^[0-9a-zA-Z]' = 0
```

Oracle

```
1 select data
2 from V
3 where translate(lower(data), '0123456789abcdefghijklmnopqrstuvwxyz',rpad('a',36,'a'))
4           = rpad('a',length(data),'a')
```



1.7 提取姓名的首字母

- 问题：你想把姓名变成首字母形式，比如LeBron James, 就可以变成L.J.

MySQL

```
1 select case
2     when cnt = 2 then
3         trim(trailing '.' from
4             concat_ws('.',
5                 substr(substring_index(name, ' ', 1), 1, 1),
6                 substr(name,
7                     length(substring_index(name, ' ', 1)) + 2, 1),
8                 substr(substring_index(name, ' ', -1), 1, 1),
9                 '.'))
10     else
11         trim(trailing '.' from
12             concat_ws('.',
13                 substr(substring_index(name, ' ', 1), 1, 1),
14                 substr(substring_index(name, ' ', -1), 1, 1)
15             ))
16     end as initials
17 from (
18 select name, length(name) - length(replace(name, ' ')) as cnt
19 from (
20 select replace('Stewie Griffin', ' ', '') as name from t1
21 ) y
22 ) x
```

Oracle and PostgreSQL

```
1 select replace(
2     replace(
3         translate(replace('Stewie Griffin', ' ', ''),
4             'abcdefghijklmnopqrstuvwxyz',
5             rpad('#', 26, '#') ), '#', '' ), ' ' || ' '
6 from t1
```



还有些字符串操作比较复杂

- 创建分隔列表、字段内部排序、解析IP地址等等
- 但一般很少使用，我们可以通过更好的设计避免出现这样的问题
- 但，SQL处理字符串的能力是非常弱小的，每个数据库都有自己的内置函数，且无法通用



思考题

- 把行数据变成以某种符号分割符的列表，比如逗号

DEPTNO EMPS

10 CLARK
10 KING
10 MILLER
20 SMITH
20 ADAMS
20 FORD
20 SCOTT
20 JONES
30 ALLEN
30 BLAKE
30 MARTIN
30 JAMES
30 TURNER
30 WARD

DEPTNO EMPS

10 CLARK,KING,MILLER
20 SMITH,JONES,SCOTT,ADAMS,FORD
30 ALLEN,WARD,MARTIN,BLAKE,TURNER,JAMES

MySQL 的内置函数GROUP_CONCAT

Oracle的内置函数SYS_CONNECT_BY_PATH



End

下一讲，我们来讲数值处理

