



# SQL 知识清单简略版

## 0 前置基础

### 0.1 基本数据类型

int, bigint, float, double, string等

### 0.2 注释

单行注释：以 -- 开头的行，将被视为注释，不被执行。

多行注释：以 / 开头，并且以 / 结尾的内容（中间可以包含多行），将被视为多行注释，不被执行

### 0.3 关键字

查询常用：select, from, where, group, order, having, join, limit, cast, as, by, with, and, or, like, in, unin, all, distinct, sort ;

创建、写入、修改、删除常用：create, exists, comment, alter, drop, inert, overwrite, into ;

常用函数：concat\_ws, explode, lateral view, desc, asc, avg, sum, if, nvl, case, when, then, else, round, over, partition, distribute, rand ;

其他：use, show, set...

## 1 表管理

### 1.1 建表

**create table if not exists** table\_name(

column1\_name column1\_type comment 'column1 note',

column2\_name column2\_type comment 'column2 note'

) **partition by** (partition\_name)  
**stored as parquet**;

## 1.2 删表

**drop table if exists** table\_name;

## 1.3 改表

关键字：**alter**

### 1.3.1 表级别

### 1.3.2 列级别

## 2 查询

### 2.1 简单查询

**select**

1+1,

8/3

注：SQL可以当作计算器使用

### 2.2 单表查询

**select**

column1,

**cast**(column2 **as** column2\_new\_type) **as** column2\_new\_name,

column3 **as** column3\_other\_name

**from** table\_name **as** table\_other\_name

注：select关键字是必须的，from、where等部分可以没有，column2中的第一个as是改变列的类型，第二个as是取别名，column3和table\_name中的as均是取别名。

## 2.3 条件查询

**select**

column1,

column2,

column3

**from** table\_name

**where** condition1 **and** condition2 **or** condition3

**order by** column1 **desc**

**limit** 100

注：where约束了返回结果需要满足的条件；order by指定了按照column1排序，desc指定降序排列，asc可以指定升序排列；limit限制返回结果的条数为100

## 2.4 分组查询

### 2.4.1 基础分组查询

**select**

**count**(\*) **as** cnt,

**avg**(column1) **as** column1\_avg,

**sum**(column2) **as** column2\_sum

column3

**from** table\_name

**group by** column3

注：group by指定按照column3的取值进行分组，count(\*)统计每个组中的数据个数，avg计算组均值，sum计算组数值之和；使用group by时，可以同时指定多个列；使用group by时，没有被用于分组的列，必须进行聚合运算，如avg、sum等。

## 2.4.2 分组前过滤

**select**

**count(\*) as** cnt,

**avg**(column1) **as** column1\_avg,

**sum**(column2) **as** column2\_sum

column3

**from** table\_name

**where** condition

**group by** column3

## 2.4.3 分组前后过滤

**select**

**count(\*) as** cnt,

**avg**(column1) **as** column1\_avg,

**sum**(column2) **as** column2\_sum

column3

**from** table\_name

**group by** column3

**having avg**(column1)>0

注：分组前过滤关键字为where,分组后过滤关键字为having。

## 2.5 多表查询

**select**

column1,

column2,

column3,

column4

**from**(

**select**

column1,

column2

**from** table1) other\_name1

**join**(

***select***

column3,

column4

***from*** table2) other\_name2

***on*** join\_condition1

***and*** join\_condition2

注：多表连接方式：inner join（join默认为inner join），outer join，left join，right join，cross join。

## 2.6 合并

(***select***

***distinct*** column1,

column2

***from*** table1\_name

***where*** condition1)

union all

(***select***

***distinct*** column1,

column2

***from*** table1\_name

**where** condition2)

## 3 高级用法

### 3.1 with 定义临时函数

--定义查询函数

**with** preprocess **as**(

**select**

column1,

**if**(column2='0', 1, 0) **as** is\_0,

**nvl**(column3) **as** not\_null,

**round**(column4, 3) **as** column4,

**case**

**when** column5 = 'str1' then 1,

**when** column5 = 'str2' then 2,

**when** column5 = 'str3' then 3,

**else** 0

**end as** column5

**from** table1\_name

), infomation **as**(

**select**

column1,

column6,

column7,

column8

**from** table2\_name

)

--使用查询函数，查询语句应出现在定义的查询函数之后

**select**

a.column1,

is\_0,

not\_null,

column4,

column5,

column6,

column7,

column8,

**concat\_ws**('-',column6,column7,column8) **as** multi\_column\_joined



**from** (

**select** \*

**from** preprocess) a

**left join**(

**select** \*

**from** information) b

**on** a.column1=b.column1

注：在调用定义的临时函数时，必须使用别名

## 3.2 窗口函数

由over partition order by等关键字构成

## 3.3 随机抽样

**select**

column1,

column2,

column3

**from** table\_name

**where** rand() < threshold

**distribute by rand()**

***sort by rand()***

***limit 10000***