

文章

[姚鑫](#) · 九月 29, 2021 阅读大约需 9 分钟

第二十九章 SQL命令 DISTINCT

第二十九章 SQL命令 DISTINCT

指定仅返回不同值的SELECT子句。

大纲

```
SELECT [DISTINCT [BY (item {,item2})] ] | [ALL]  
select-item {,select-item2}
```

参数

- DISTINCT - 可选-返回组合选择项值唯一的行。
- DISTINCT BY (item {,item2}) - 可选-返回按(项)值唯一的行的选择项值。
- ALL - 可选-返回结果集中的所有行。默认设置。

描述

可选DISTINCT子句出现在SELECT关键字之后、可选TOP子句和第一个SELECT-ITEM之前。

DISTINCT子句应用于SELECT语句的结果集。它将每个不同(唯一)值返回的行数限制为一个任意行。如果未指定DISTINCT子句，则默认情况下显示满足选择条件的所有行。ALL子句与不指定DEFAULT子句相同；如果指定ALL，SELECT将返回表中满足选择条件的所有行。

DISTINCT从句有两种形式：

- SELECT DISTINCT：为选择项值的每个唯一组合返回一行。可以指定一个或多个选择项。例如，以下查询返回一行，其中包含HomeState和Age值的每个唯一组合的HomeState和Age值：

```
SELECT DISTINCT Home_State, Age FROM Sample.Person
```

- SELECT DISTINCT BY(item)：为项目值的每个唯一组合返回一行。可以指定单个项目或逗号分隔的项目列表。指定的项目或项目列表必须用括号括起来。可以在by关键字和圆括号之间指定或省略空格。选择项列表可以(但不一定)包括指定的项。例如，以下查询返回一行，其中包含HomeState和Age值的每个唯一组合的Name和Age值：

```
SELECT DISTINCT BY (Home_State, Age) Name, Age FROM Sample.Person
```

项目字段必须按列名指定。有效值包括以下值：列名(DISTINCT BY(City))；%ID(返回所有行)；指定列名的标量函数(DISTINCT BY(ROUND(Age, -1))；指定列名的排序函数(DISTINCT BY(%Exact(City))。不能按列别名指定字段；尝试这样做会生成SQLCODE-29错误。不能按列号指定字段；这将被解释为文字，并返回一行。将文字指定为DISTINCT子句中的项值将返回1行；返回哪行是不确定的。因此，指定7、' Chicago '、' '、0或NULL都返回1行。但是，如果将文字指定

为逗号分隔列表中的项值，则该文字将被忽略，并且DISTINCT将为指定字段名的每个唯一组合选择一行。

DISTINCT子句在TOP子句之前应用。如果两者都指定，则SELECT只返回具有唯一值的行，唯一值是在TOP子句中指定的唯一值行数。

如果DISTINCT子句中指定的列包含NULL(不包含值)行，则DISTINCT将返回一行作为DISTINCT(唯一)值的NULL，如以下示例所示：

```
SELECT DISTINCT FavoriteColors FROM Sample.Person
```

```
SELECT DISTINCT BY (FavoriteColors) Name,FavoriteColors FROM Sample.Person  
ORDER BY FavoriteColors
```

DISTINCT子句在嵌入式SQL简单查询中没有意义，因为在这种类型的嵌入式SQL中，SELECT始终只返回一行数据。但是，嵌入式SQL基于游标的查询可以返回多行数据；在基于游标的查询中，DISTINCT子句只返回唯一值行。

DISTINCT和ORDER BY

DISTINCT子句在ORDER BY子句之前应用。因此，DISTINCT和ORDER BY的组合将首先选择满足DISTINCT子句的任意行，然后根据ORDER BY子句对这些行进行排序。

DISTINCT和GROUP BY

DISTINCT和GROUP BY这两个记录按指定字段(或多个字段)分组，并为该字段的每个唯一值返回一条记录。它们之间的一个重要区别是DISTINCT在分组之前计算聚合函数。GROUP BY计算分组后的聚合函数。以下示例显示了这种差异：

```
SELECT DISTINCT BY (ROUND(Age, -1)) Age,AVG(Age) AS AvgAge FROM Sample.Person  
/* Avg(Age)????????????? */
```

```
SELECT Age,AVG(Age) AS AvgAge FROM Sample.Person GROUP BY ROUND(Age, -1)  
/* Avg(Age)????????????? */
```

DISTINCT子句可以用一个或多个聚合函数字段指定，尽管这很少有意义，因为聚合函数返回单个值。因此，下面的示例返回单行：

```
SELECT DISTINCT BY (AVG(Age)) Name, Age,AVG(Age) FROM Sample.Person
```

注意：如果将聚合函数作为唯一项或选择项的DISTINCT子句与GROUP BY子句一起使用，则DISTINCT子句将被忽略。可以使用子查询实现DISTINCT、聚合函数和GROUP BY的预期组合。

字母大小写与DISTINCT优化

根据为字段定义的排序规则类型，将字符串值不同地分组在一起。默认情况下，字符串数据类型字段使用SQLUPPER排序规则定义，该排序规则不区分大小写。

如果字段/特性排序规则类型为SQLUPPER，则分组的字段值将全部以大写字母返回。要按原始字母大小写对值进行

```
SELECT DISTINCT BY (Home_City) Name, Home_City FROM Sample.Person
/* ?Home_City?????????????????????????????????????????'new york'.
*/

SELECT DISTINCT BY (Home_City) Name,%EXACT(Home_City) FROM Sample.Person
/* ?Home_City?????????????????????????????????????(???????)?????????'New York'? 'new York'?????????
??? */

SELECT DISTINCT BY (%EXACT(Home_City)) Name, Home_City FROM Sample.Person
/* ?Home_City?????????????????????????????????????(???????)?
??? 'New York'? 'new York'?????
?????. */
```

此默认设置按字母值的大写排序规则对字母值进行分组。此优化利用选定字段的索引。因此，只有一个或多个选定字段存在索引时才有意义。它对存储在索引中的字段值进行排序；字母字符串以全部大写字母返回。您可以设置此系统范围的选项，然后使用%exact排序规则函数为特定查询覆盖它以保留字母大小写。

也可以使用`$SYSTEM.SQL.Util.SetOption()`方法快速区分选项在系统范围内设置此选项。要确定当前设置，请调用`$SYSTEM.SQL.CurrentSettings()`，它显示打开的不同优化设置；默认值为1。

- 流字段：DISTINCT对流字段的OID进行操作，而不是对其实际数据进行操作。因为所有流字段OID都是唯一值，所以DISTINCT对实际流字段重复数据值没有影响。DISTINCT BY(StreamField)将流字段为空的记录数减少到一个空记录。
- 星号语法：DISTINCT*语法是合法的，但没有意义，因为根据定义，所有行都包含一些不同的唯一标识符。不同于(*)的语法不合法。
- 子查询：在子查询中使用DISTINCT子句是合法的，但没有意义，因为子查询返回单个值。
- 未选择行数据：DISTINCT子句可以与不访问任何表数据的SELECT一起使用。如果SELECT包含FROM子句，则在一行中指定DISTINCT结果将包含这些非表值；如果未指定DISTINCT(或TOP)，则SELECT将产生与FROM子句表中的行数相同的行数。如果SELECT不包含FROM子句，则DISTINCT是合法的，但没有意义。
- 聚合函数：可以在聚合函数中使用DISTINCT子句，以仅选择要包含在聚合中的不同(唯一)字段值。与SELECT DISTINCT子句不同，聚合函数中的DISTINCT不包括NULL作为DISTINCT(唯一)值。请注意，MAX和MIN聚合函数分析DISTINCT子句语法没有错误，但此语法不执行任何操作。

指定DISTINCT关键字会导致基于游标的嵌入式SQL查询不设置%ROWID变量。即使DISTINCT不限制返回的行数，也不设置%ROWID。下面的示例显示了这一点：

```
ClassMethod Distinct()
{
    s %ROWID = 999
    &sql(
        DECLARE EmpCursor3 CURSOR FOR
```

```
        SELECT DISTINCT Name, Home_State
        INTO :name,:state FROM Sample.Person
        WHERE Home_State %STARTSWITH 'M'
    )
    &sql(
        OPEN EmpCursor3
    )
    q:(SQLCODE '= 0)
    for {
        &sql(
            FETCH EmpCursor3
        )
        q:SQLCODE
        w !,"RowID: ",%ROWID," row count: ",%ROWCOUNT
        w " Name=",name," State=",state
    }
    &sql(
        CLOSE EmpCursor3
    )
}
```

查询行为的这种更改仅适用于基于游标的嵌入式SQL SELECT查询。动态SQL SELECT查询和非游标嵌入式SQL SELECT查询从未设置%ROWID。

DISTINCT和事务处理

指定DISTINCT关键字会导致查询检索所有当前数据，包括当前事务尚未提交的数据。忽略事务的READ COMMITTED隔离模式参数(如果设置)；在READ UNCOMMITTED模式下检索所有数据。

示例

以下查询为每个不同的HomeState值返回一行：

```
SELECT DISTINCT Home_State FROM Sample.Person
ORDER BY Home_State
```

以下查询为每个不同的HomeState值返回一行，但返回该行的其他字段。无法预测检索到的是哪一行：

```
SELECT DISTINCT BY (Home_State) %ID,Name,Home_State,Office_State FROM Sample.Person
ORDER BY Home_State
```

以下查询为HomeState和OfficeState值的每个不同组合返回一行。根据数据的不同，它要么返回更多行，要么返回与上一个示例相同的行数：

```
SELECT DISTINCT BY (Home_State,Office_State) %ID,Name,Home_State,Office_State FROM Sample.Person
ORDER BY Home_State,Office_State
```

以下查询使用DISTINCT BY为每个不同的名称长度返回一行：

```
SELECT DISTINCT BY ($LENGTH(Name)) Name,$LENGTH(Name) AS lname
FROM Sample.Person
ORDER BY lname
```

下面的查询使用DISTINCT BY为FavoriteColors
%List值的每个不同的第一个元素返回一行。它列出FavoriteColors为空的各行：

```
SELECT DISTINCT BY ($LIST(FavoriteColors,1)) Name,FavoriteColors,$LIST(FavoriteColors
,1) AS FirstColor
FROM Sample.Person
```

以下查询按排序规则升序返回从Sample.Person检索到的前20个不同的Home_State值。“top”行反映Sample.Person中所有行的ORDER BY子句排序。

```
SELECT DISTINCT TOP 20 Home_State FROM Sample.Person ORDER BY Home_State
```

以下查询在主查询和WHERE子句子查询中都使用DISTINCT。它返回Sample.Person中的前20个不同的Home_State值，这些值也在Sample.Employee中。如果未提供子查询DISTINCT，它将检索Sample.Person中与Sample.Employee中随机选择的Home_State值匹配的DISTINCT Home_State值：

```
SELECT DISTINCT TOP 20 Home_State FROM Sample.Person
WHERE Home_State IN(SELECT DISTINCT TOP 20 Home_State FROM Sample.Employee)
ORDER BY Home_State
```

以下查询返回前20个不同的FavoriteColors值。这反映了Sample.Person中所有行的ORDER BY子句排序。众所周知，FavoriteColors字段具有NULL，因此FavoriteColors为NULL的各行出现在排序规则序列的顶部。

```
SELECT DISTINCT BY (FavoriteColors) TOP 20 FavoriteColors,Name FROM Sample.Person
ORDER BY FavoriteColors
```

还要注意，在前面的示例中，因为FavoriteColors是一个列表字段，所以归类序列包括元素长度字节。因此，以三个字母元素(红色)开头的不同列表值列在以四个字母元素(蓝色)开头的列表值之前。

[#SQL #Caché](#)

源

URL:

<https://cn.community.intersystems.com/post/%E7%AC%AC%E4%BA%8C%E5%8D%81%E4%B9%9D%E7%AB%A0-sql%E5%91%BD%E4%BB%A4-distinct>