15.2.20 WITH (Common Table Expressions)

A common table expression (CTE) is a named temporary result set that exists within the scope of a single statement and that can be referred to later within that statement, possibly multiple times. The following discussion describes how to write statements that use CTEs.

- Common Table Expressions
- Recursive Common Table Expressions
- Limiting Common Table Expression Recursion
- Recursive Common Table Expression Examples
- Common Table Expressions Compared to Similar Constructs

For information about CTE optimization, see Section 10.2.2.4, "Optimizing Derived Tables, View References, and Common Table Expressions with Merging or Materialization".

Additional Resources

These articles contain additional information about using CTEs in MySQL, including many examples:

- MySQL 8.0 Labs: [Recursive] Common Table Expressions in MySQL (CTEs)
- MySQL 8.0 Labs: [Recursive] Common Table Expressions in MySQL (CTEs), Part Two how to generate series
- MySQL 8.0 Labs: [Recursive] Common Table Expressions in MySQL (CTEs), Part Three hierarchies
- MySQL 8.0.1: [Recursive] Common Table Expressions in MySQL (CTEs), Part Four depth-first or breadth-first traversal, transitive closure, cycle avoidance

Common Table Expressions

To specify common table expressions, use a $\underline{\text{WITH}}$ clause that has one or more comma-separated subclauses. Each subclause provides a subquery that produces a result set, and associates a name with the subquery. The following example defines CTEs named <code>cte1</code> and <code>cte2</code> in the $\underline{\text{WITH}}$ clause, and refers to them in the top-level <code>SELECT</code> that follows the <code>WITH</code> clause:

```
WITH

cte1 AS (SELECT a, b FROM table1),

cte2 AS (SELECT c, d FROM table2)

SELECT b, d FROM cte1 JOIN cte2

WHERE cte1.a = cte2.c;
```

In the statement containing the $\underline{\text{WITH}}$ clause, each CTE name can be referenced to access the corresponding CTE result set.

A CTE name can be referenced in other CTEs, enabling CTEs to be defined based on other CTEs.

A CTE can refer to itself to define a recursive CTE. Common applications of recursive CTEs include series generation and traversal of hierarchical or tree-structured data.

Common table expressions are an optional part of the syntax for DML statements. They are defined using a WITH clause:

```
with_clause:
    WITH [RECURSIVE]
    cte_name [(col_name [, col_name] ...)] AS (subquery)
    [, cte_name [(col_name [, col_name] ...)] AS (subquery)] ...
```

 cte_name names a single common table expression and can be used as a table reference in the statement containing the with clause.

The *subquery* part of AS (*subquery*) is called the "subquery of the CTE" and is what produces the CTE result set. The parentheses following AS are required.

A common table expression is recursive if its subquery refers to its own name. The RECURSIVE keyword must be included if any CTE in the $\underline{\text{WITH}}$ clause is recursive. For more information, see Recursive Common Table Expressions.

Determination of column names for a given CTE occurs as follows:

• If a parenthesized list of names follows the CTE name, those names are the column names:

```
WITH cte (col1, col2) AS
(
SELECT 1, 2
UNION ALL
SELECT 3, 4
```

```
)
SELECT col1, col2 FROM cte;
```

The number of names in the list must be the same as the number of columns in the result set.

• Otherwise, the column names come from the select list of the first <u>SELECT</u> within the AS (*subquery*) part:

```
WITH cte AS
(
SELECT 1 AS col1, 2 AS col2
UNION ALL
SELECT 3, 4
)
SELECT col1, col2 FROM cte;
```

A WITH clause is permitted in these contexts:

• At the beginning of SELECT, UPDATE, and DELETE statements.

```
WITH ... SELECT ...
WITH ... UPDATE ...
WITH ... DELETE ...
```

• At the beginning of subqueries (including derived table subqueries):

```
SELECT ... WHERE id IN (WITH ... SELECT ...) ...
SELECT * FROM (WITH ... SELECT ...) AS dt ...
```

• Immediately preceding SELECT for statements that include a SELECT statement:

```
INSERT ... WITH ... SELECT ...
REPLACE ... WITH ... SELECT ...
CREATE TABLE ... WITH ... SELECT ...
CREATE VIEW ... WITH ... SELECT ...
DECLARE CURSOR ... WITH ... SELECT ...
EXPLAIN ... WITH ... SELECT ...
```

Only one $\underline{\text{WITH}}$ clause is permitted at the same level. $\underline{\text{WITH}}$ followed by $\underline{\text{WITH}}$ at the same level is not permitted, so this is illegal:

```
WITH cte1 AS (...) WITH cte2 AS (...) SELECT ...
```

To make the statement legal, use a single WITH clause that separates the subclauses by a comma:

```
WITH cte1 AS (...), cte2 AS (...) SELECT ...
```

However, a statement can contain multiple WITH clauses if they occur at different levels:

```
WITH cte1 AS (SELECT 1)
SELECT * FROM (WITH cte2 AS (SELECT 2) SELECT * FROM cte2 JOIN cte1) AS dt;
```

A <u>WITH</u> clause can define one or more common table expressions, but each CTE name must be unique to the clause. This is illegal:

```
WITH cte1 AS (...), cte1 AS (...) SELECT ...
```

To make the statement legal, define the CTEs with unique names:

```
WITH cte1 AS (...), cte2 AS (...) SELECT ...
```

A CTE can refer to itself or to other CTEs:

- A self-referencing CTE is recursive.
- A CTE can refer to CTEs defined earlier in the same WITH clause, but not those defined later.

This constraint rules out mutually-recursive CTEs, where <code>ctel</code> references <code>ctel</code> and <code>ctel</code> references <code>ctel</code>. One of those references must be to a CTE defined later, which is not permitted.

• A CTE in a given query block can refer to CTEs defined in query blocks at a more outer level, but not CTEs defined in query blocks at a more inner level.

For resolving references to objects with the same names, derived tables hide CTEs; and CTEs hide base tables, TEMPORARY tables, and views. Name resolution occurs by searching for objects in the same query block, then proceeding to outer blocks in turn while no object with the name is found.

Like derived tables, a CTE cannot contain outer references prior to MySQL 8.0.14. This is a MySQL restriction that is lifted in MySQL 8.0.14, not a restriction of the SQL standard. For additional syntax considerations specific to recursive CTEs, see Recursive Common Table Expressions.

Recursive Common Table Expressions

A recursive common table expression is one having a subquery that refers to its own name. For example:

```
WITH RECURSIVE cte (n) AS

(
    SELECT 1
    UNION ALL
    SELECT n + 1 FROM cte WHERE n < 5
)

SELECT * FROM cte;
```

When executed, the statement produces this result, a single column containing a simple linear sequence:

A recursive CTE has this structure:

• The WITH clause must begin with WITH RECURSIVE if any CTE in the WITH clause refers to itself. (If no CTE refers to itself, RECURSIVE is permitted but not required.)

If you forget RECURSIVE for a recursive CTE, this error is a likely result:

```
ERROR 1146 (42SO2): Table 'cte_name' doesn't exist
```

• The recursive CTE subquery has two parts, separated by UNION ALL or UNION [DISTINCT]:

```
SELECT ... -- return initial row set
UNION ALL
SELECT ... -- return additional row sets
```

The first <u>SELECT</u> produces the initial row or rows for the CTE and does not refer to the CTE name. The second <u>SELECT</u> produces additional rows and recurses by referring to the CTE name in its FROM clause. Recursion ends when this part produces no new rows. Thus, a recursive CTE consists of a nonrecursive SELECT part followed by a recursive SELECT part.

Each SELECT part can itself be a union of multiple SELECT statements.

- The types of the CTE result columns are inferred from the column types of the nonrecursive <u>SELECT</u> part only, and the columns are all nullable. For type determination, the recursive <u>SELECT</u> part is ignored.
- If the nonrecursive and recursive parts are separated by <u>UNION DISTINCT</u>, duplicate rows are eliminated. This is useful for queries that perform transitive closures, to avoid infinite loops.
- Each iteration of the recursive part operates only on the rows produced by the previous iteration. If the recursive part has multiple query blocks, iterations of each query block are scheduled in unspecified order, and each query block operates on rows that have been produced either by its previous iteration or by other query blocks since that previous iteration's end.

The recursive CTE subquery shown earlier has this nonrecursive part that retrieves a single row to produce the initial row set:

```
SELECT 1
```

The CTE subquery also has this recursive part:

```
SELECT n + 1 FROM cte WHERE n < 5
```

At each iteration, that $\underline{\texttt{SELECT}}$ produces a row with a new value one greater than the value of n from the previous row set. The first iteration operates on the initial row set (1) and produces 1+1=2; the second iteration operates on the first iteration's row set (2) and produces 2+1=3; and so forth. This continues until recursion ends, which occurs when n is no longer less than 5.

If the recursive part of a CTE produces wider values for a column than the nonrecursive part, it may be necessary to widen the column in the nonrecursive part to avoid data truncation. Consider this statement:

```
WITH RECURSIVE cte AS
(
   SELECT 1 AS n, 'abc' AS str
   UNION ALL
   SELECT n + 1, CONCAT(str, str) FROM cte WHERE n < 3
)
SELECT * FROM cte;</pre>
```

In nonstrict SQL mode, the statement produces this output:

The str column values are all 'abc' because the nonrecursive <u>SELECT</u> determines the column widths. Consequently, the wider str values produced by the recursive SELECT are truncated.

In strict SQL mode, the statement produces an error:

```
ERROR 1406 (22001): Data too long for column 'str' at row 1
```

To address this issue, so that the statement does not produce truncation or errors, use $\underline{\mathtt{CAST}()}$ in the nonrecursive SELECT to make the \mathtt{str} column wider:

```
WITH RECURSIVE cte AS
(
   SELECT 1 AS n, CAST('abc' AS CHAR(20)) AS str
   UNION ALL
   SELECT n + 1, CONCAT(str, str) FROM cte WHERE n < 3
)
SELECT * FROM cte;</pre>
```

Now the statement produces this result, without truncation:

Columns are accessed by name, not position, which means that columns in the recursive part can access columns in the nonrecursive part that have a different position, as this CTE illustrates:

```
WITH RECURSIVE cte AS

(
SELECT 1 AS n, 1 AS p, -1 AS q
UNION ALL
SELECT n + 1, q * 2, p * 2 FROM cte WHERE n < 5
)
SELECT * FROM cte;
```

Because p in one row is derived from q in the previous row, and vice versa, the positive and negative values swap positions in each successive row of the output:

Some syntax constraints apply within recursive CTE subqueries:

- The recursive SELECT part must not contain these constructs:
 - Aggregate functions such as SUM()
 - Window functions
 - GROUP BY

- ORDER BY
- DISTINCT

Prior to MySQL 8.0.19, the recursive <u>SELECT</u> part of a recursive CTE also could not use a LIMIT clause. This restriction is lifted in MySQL 8.0.19, and LIMIT is now supported in such cases, along with an optional OFFSET clause. The effect on the result set is the same as when using LIMIT in the outermost SELECT, but is also more efficient, since using it with the recursive SELECT stops the generation of rows as soon as the requested number of them has been produced.

These constraints do not apply to the nonrecursive <u>SELECT</u> part of a recursive CTE. The prohibition on DISTINCT applies only to UNION members; UNION DISTINCT is permitted.

• The recursive <u>SELECT</u> part must reference the CTE only once and only in its FROM clause, not in any subquery. It can reference tables other than the CTE and join them with the CTE. If used in a join like this, the CTE must not be on the right side of a LEFT JOIN.

These constraints come from the SQL standard, other than the MySQL-specific exclusions of ORDER BY, LIMIT (MySQL 8.0.18 and earlier), and DISTINCT.

For recursive CTEs, <u>EXPLAIN</u> output rows for recursive <u>SELECT</u> parts display Recursive in the Extra column.

Cost estimates displayed by <u>EXPLAIN</u> represent cost per iteration, which might differ considerably from total cost. The optimizer cannot predict the number of iterations because it cannot predict at what point the WHERE clause becomes false.

CTE actual cost may also be affected by result set size. A CTE that produces many rows may require an internal temporary table large enough to be converted from in-memory to on-disk format and may suffer a performance penalty. If so, increasing the permitted in-memory temporary table size may improve performance; see Section 10.4.4, "Internal Temporary Table Use in MySQL".

Limiting Common Table Expression Recursion

It is important for recursive CTEs that the recursive $\underline{\mathtt{SELECT}}$ part include a condition to terminate recursion. As a development technique to guard against a runaway recursive CTE, you can force termination by placing a limit on execution time:

• The cte_max_recursion_depth system variable enforces a limit on the number of recursion levels for CTEs. The server terminates execution of any CTE that recurses more levels than the value of this variable.

- The max_execution_time system variable enforces an execution timeout for SELECT statements executed within the current session.
- The MAX_EXECUTION_TIME optimizer hint enforces a per-query execution timeout for the SELECT statement in which it appears.

Suppose that a recursive CTE is mistakenly written with no recursion execution termination condition:

```
WITH RECURSIVE cte (n) AS

(
SELECT 1
UNION ALL
SELECT n + 1 FROM cte
)
SELECT * FROM cte;
```

By default, cte_max_recursion_depth has a value of 1000, causing the CTE to terminate when it recurses past 1000 levels. Applications can change the session value to adjust for their requirements:

```
SET SESSION cte_max_recursion_depth = 10; -- permit only shallow recursion SET SESSION cte_max_recursion_depth = 1000000; -- permit deeper recursion
```

You can also set the global <u>cte_max_recursion_depth</u> value to affect all sessions that begin subsequently.

For queries that execute and thus recurse slowly or in contexts for which there is reason to set the cte_max_recursion_depth value very high, another way to guard against deep recursion is to set a persession timeout. To do so, execute a statement like this prior to executing the CTE statement:

```
SET max_execution_time = 1000; -- impose one second timeout
```

Alternatively, include an optimizer hint within the CTE statement itself:

```
WITH RECURSIVE cte (n) AS
(
    SELECT 1
    UNION ALL
    SELECT n + 1 FROM cte
)
SELECT /*+ SET_VAR(cte_max_recursion_depth = 1M) */ * FROM cte;
```

```
WITH RECURSIVE cte (n) AS
(
    SELECT 1
    UNION ALL
    SELECT n + 1 FROM cte
)
SELECT /*+ MAX_EXECUTION_TIME(1000) */ * FROM cte;
```

Beginning with MySQL 8.0.19, you can also use LIMIT within the recursive query to impose a maximum number of rows to be returned to the outermost SELECT, for example:

```
WITH RECURSIVE cte (n) AS

(
SELECT 1
UNION ALL
SELECT n + 1 FROM cte LIMIT 10000
)

SELECT * FROM cte;
```

You can do this in addition to or instead of setting a time limit. Thus, the following CTE terminates after returning ten thousand rows or running for one second (1000 milliseconds), whichever occurs first:

```
WITH RECURSIVE cte (n) AS

(
SELECT 1
UNION ALL
SELECT n + 1 FROM cte LIMIT 10000
)

SELECT /*+ MAX_EXECUTION_TIME(1000) */ * FROM cte;
```

If a recursive query without an execution time limit enters an infinite loop, you can terminate it from another session using KILL QUERY. Within the session itself, the client program used to run the query might provide a way to kill the query. For example, in **mysql**, typing **Control+C** interrupts the current statement.

Recursive Common Table Expression Examples

As mentioned previously, recursive common table expressions (CTEs) are frequently used for series generation and traversing hierarchical or tree-structured data. This section shows some simple examples of these techniques.

- Fibonacci Series Generation
- Date Series Generation
- Hierarchical Data Traversal

Fibonacci Series Generation

A Fibonacci series begins with the two numbers 0 and 1 (or 1 and 1) and each number after that is the sum of the previous two numbers. A recursive common table expression can generate a Fibonacci series if each row produced by the recursive <u>SELECT</u> has access to the two previous numbers from the series. The following CTE generates a 10-number series using 0 and 1 as the first two numbers:

```
WITH RECURSIVE fibonacci (n, fib_n, next_fib_n) AS
(
   SELECT 1, 0, 1
   UNION ALL
   SELECT n + 1, next_fib_n, fib_n + next_fib_n
     FROM fibonacci WHERE n < 10
)
SELECT * FROM fibonacci;</pre>
```

The CTE produces this result:

```
+----+
| n | fib_n | next_fib_n |
+----+
  1 | 0 |
  2 |
      1 |
              1 |
  3 |
      1 |
              2 |
  4 |
      2 |
              3 |
  5 |
      3 |
              5 |
  6 |
      5 |
              8 |
  7 |
      8 |
              13 |
  8 | 13 |
              21 |
 9 |
      21 |
              34 |
  10 |
      34 |
              55 |
+----+
```

How the CTE works:

• n is a display column to indicate that the row contains the n-th Fibonacci number. For example, the 8th Fibonacci number is 13.

- The fib n column displays Fibonacci number n.
- The next_fib_n column displays the next Fibonacci number after number n. This column provides the next series value to the next row, so that row can produce the sum of the two previous series values in its fib_n column.
- Recursion ends when n reaches 10. This is an arbitrary choice, to limit the output to a small set of rows.

The preceding output shows the entire CTE result. To select just part of it, add an appropriate WHERE clause to the top-level SELECT. For example, to select the 8th Fibonacci number, do this:

```
mysql> WITH RECURSIVE fibonacci ...

SELECT fib_n FROM fibonacci WHERE n = 8;
+----+
| fib_n |
+----+
| 13 |
+----+
```

Date Series Generation

A common table expression can generate a series of successive dates, which is useful for generating summaries that include a row for all dates in the series, including dates not represented in the summarized data.

Suppose that a table of sales numbers contains these rows:

This guery summarizes the sales per day:

However, that result contains "holes" for dates not represented in the range of dates spanned by the table. A result that represents all dates in the range can be produced using a recursive CTE to generate that set of dates, joined with a LEFT JOIN to the sales data.

Here is the CTE to generate the date range series:

```
WITH RECURSIVE dates (date) AS
(
   SELECT MIN(date) FROM sales
   UNION ALL
   SELECT date + INTERVAL 1 DAY FROM dates
   WHERE date + INTERVAL 1 DAY <= (SELECT MAX(date) FROM sales)
)
SELECT * FROM dates;</pre>
```

The CTE produces this result:

How the CTF works:

- The nonrecursive SELECT produces the lowest date in the date range spanned by the sales table.
- Each row produced by the recursive SELECT adds one day to the date produced by the previous row.
- Recursion ends after the dates reach the highest date in the date range spanned by the sales table.

Joining the CTE with a LEFT JOIN against the sales table produces the sales summary with a row for each date in the range:

```
WITH RECURSIVE dates (date) AS

(
SELECT MIN(date) FROM sales
UNION ALL
SELECT date + INTERVAL 1 DAY FROM dates
WHERE date + INTERVAL 1 DAY <= (SELECT MAX(date) FROM sales)
)
SELECT dates.date, COALESCE(SUM(price), 0) AS sum_price
FROM dates LEFT JOIN sales ON dates.date = sales.date
GROUP BY dates.date
ORDER BY dates.date;
```

The output looks like this:

```
+----+
| date | sum_price |
+----+
| 2017-01-03 | 300.00 |
| 2017-01-04 |
            0.00 |
| 2017-01-05 |
             0.00
| 2017-01-06 | 50.00 |
| 2017-01-07 |
             0.00
| 2017-01-08 | 180.00 |
| 2017-01-09 |
            0.00
| 2017-01-10 |
             5.00 |
+----+
```

Some points to note:

• Are the queries inefficient, particularly the one with the $\underline{\text{MAX}()}$ subquery executed for each row in the recursive $\underline{\text{SELECT}}$? $\underline{\text{EXPLAIN}}$ shows that the subquery containing $\underline{\text{MAX}()}$ is evaluated only once and the result is cached.

• The use of COALESCE() avoids displaying NULL in the sum_price column on days for which no sales
data occur in the sales table.

Hierarchical Data Traversal

Recursive common table expressions are useful for traversing data that forms a hierarchy. Consider these statements that create a small data set that shows, for each employee in a company, the employee name and ID number, and the ID of the employee's manager. The top-level employee (the CEO), has a manager ID of NULL (no manager).

```
CREATE TABLE employees (
  id
           INT PRIMARY KEY NOT NULL,
 name
          VARCHAR(100) NOT NULL,
 manager_id INT NULL,
 INDEX (manager_id),
FOREIGN KEY (manager_id) REFERENCES employees (id)
);
INSERT INTO employees VALUES
(333, "Yasmina", NULL), # Yasmina is the CEO (manager_id is NULL)
(198, "John", 333),
                      # John has ID 198 and reports to 333 (Yasmina)
(692, "Tarek", 333),
(29, "Pedro", 198),
(4610, "Sarah", 29),
(72, "Pierre", 29),
(123, "Adil", 692);
```

The resulting data set looks like this:

```
mysql> SELECT * FROM employees ORDER BY id;
+----+
| id | name | manager_id |
+----+
 29 | Pedro |
                198 |
 72 | Pierre |
                 29
| 123 | Adil |
                692 |
| 198 | John
                333 |
| 333 | Yasmina |
               NULL |
| 692 | Tarek |
                333 |
| 4610 | Sarah |
                 29 |
+----+
```

To produce the organizational chart with the management chain for each employee (that is, the path from CEO to employee), use a recursive CTE:

```
WITH RECURSIVE employee_paths (id, name, path) AS
(
    SELECT id, name, CAST(id AS CHAR(200))
    FROM employees
    WHERE manager_id IS NULL
    UNION ALL
    SELECT e.id, e.name, CONCAT(ep.path, ',', e.id)
    FROM employee_paths AS ep JOIN employees AS e
        ON ep.id = e.manager_id
)
SELECT * FROM employee_paths ORDER BY path;
```

The CTE produces this output:

How the CTE works:

• The nonrecursive SELECT produces the row for the CEO (the row with a NULL manager ID).

The path column is widened to CHAR (200) to ensure that there is room for the longer path values produced by the recursive SELECT.

- Each row produced by the recursive <u>SELECT</u> finds all employees who report directly to an employee produced by a previous row. For each such employee, the row includes the employee ID and name, and the employee management chain. The chain is the manager's chain, with the employee ID added to the end.
- Recursion ends when employees have no others who report to them.

To find the path for a specific employee or employees, add a WHERE clause to the top-level <u>SELECT</u>. For example, to display the results for Tarek and Sarah, modify that <u>SELECT</u> like this:

Common Table Expressions Compared to Similar Constructs

Common table expressions (CTEs) are similar to derived tables in some ways:

- Both constructs are named.
- Both constructs exist for the scope of a single statement.

Because of these similarities, CTEs and derived tables often can be used interchangeably. As a trivial example, these statements are equivalent:

```
WITH cte AS (SELECT 1) SELECT * FROM cte;
SELECT * FROM (SELECT 1) AS dt;
```

However, CTEs have some advantages over derived tables:

- A derived table can be referenced only a single time within a query. A CTE can be referenced multiple times. To use multiple instances of a derived table result, you must derive the result multiple times.
- A CTE can be self-referencing (recursive).
- One CTF can refer to another.
- A CTE may be easier to read when its definition appears at the beginning of the statement rather than embedded within it.

CTEs are similar to tables created with CREATE [TEMPORARY] TABLE but need not be defined or dropped explicitly. For a CTE, you need no privileges to create tables.

© 2024 Oracle