Table of Contents

My SQL has a RollUp phrase that you can use with the Group By clause. The purpose of the roll up is to get grand totals of the result set. The Grand total is the total of all of the individual rows. Compare the output of the first two queries below. The first groups by the department id and has one row per department. The second adds the roll up and gets an additional row that is the total values for all of the departments. This is called the grand total super aggregate row.

As with all aggregate queries you need to pay attention to the Grouping expressions and use using the correct aggregate.

1. Group by WITH ROLLUP

The Group by With RollUp clause is part of MySQL. We will start with a regular group by query.

Demo 01: Do a group by each department with aggregates on salary and emp_id and we get one row for each department.

```
Select dept_id, sum(salary), count(salary), count(emp_id)
From a_testbed.adv_emp
Group by dept_id;
+---------+-------------+---------------+---------------+
| dept_id | SUM(salary) | COUNT(salary) | COUNT(emp_id) |
+---------+-------------+---------------+---------------+
|      10 |       45000 |             1 |             1 |
|      20 |      150900 |             6 |             6 |
|      30 |      153000 |             5 |             5 |
|      45 |      305000 |             9 |             9 |
+---------+-------------+---------------+---------------+
4 rows in set (0.00 sec)
```

Demo 02: Adding the `WITH ROLLUP` phrase. This gives us one additional row which totals the various groups. The first column in this row has a null to signify that this is a rollup row; there is no valid entry for a department id when the row refers to all the departments.

```
Select dept_id, sum(salary), count(salary), count(emp_id)
From a_testbed.adv_emp
Group by dept_id WITH ROLLUP;
+---------+-------------+---------------+---------------+
| dept_id | SUM(salary) | COUNT(salary) | COUNT(emp_id) |
+---------+-------------+---------------+---------------+
|      10 |       45000 |             1 |             1 |
|      20 |      150900 |             6 |             6 |
|      30 |      153000 |             5 |             5 |
|      45 |      305000 |             9 |             9 |
|    NULL |      653900 |            21 |            21 |
+---------+-------------+---------------+---------------+
5 rows in set (0.00 sec)
```

Demo 03: We can use coalesce to improve the output. But we do need to be certain that the only null we can have in that column is the null that was added for the Rollup

```
Select
  concat('dept: ' , coalesce(dept_id, 'All' ) ) as Dept
, sum(salary)
, count(salary)
, count(emp_id)
From a_testbed.adv_emp
Group by dept_id WITH ROLLUP;
+-----------+-------------+---------------+---------------+
| Dept      | sum(salary) | count(salary) | count(emp_id) |
+-----------+-------------+---------------+---------------+
| dept: 10  |       45000 |             1 |             1 |
| dept: 20  |      150900 |             6 |             6 |
| dept: 30  |      153000 |             5 |             5 |
| dept: 45  |      305000 |             9 |             9 |
| dept: All |      653900 |            21 |            21 |
+-----------+-------------+---------------+---------------+
5 rows in set (0.00 sec)
```

Demo 04: We could have gotten this result with a Union query that did the totals for the table in the second query.

```
Select concat('dept: ' ,dept_id) as dept
, sum(salary), count(salary), count(emp_id)
From a_testbed.adv_emp
Group by dept_id
union
Select 'dept: All'
, sum(salary), count(salary), count(emp_id)
From a_testbed.adv_emp ;
```

It is not unusual to have more than one way to perform a task. Sometimes there is a difference in efficiency of execution; sometime there is a difference in efficiency of maintenance; sometimes it is just a matter of personal preference.

Demo 05: Doing a rollup on the salary attribute- how many employees do we have at each salary level?

```
Select salary,  count(emp_id)
From a_testbed.adv_emp
Group by salary WITH ROLLUP;
+--------+---------------+
| salary | COUNT(emp_id) |
+--------+---------------+
|  20000 |             1 |
|  22000 |             1 |
|  24000 |             1 |
|  25000 |             4 |
|  28000 |             6 |
|  30900 |             1 |
|  32000 |             2 |
|  45000 |             5 |
|   NULL |            21 |
+--------+---------------+
9 rows in set (0.00 sec)
```

We want a label for that last column- otherwise it looks like we have 21show warnings; employees with no salary.

```
Select coalesce(salary, 'all') as salaryCol,  count(emp_id)
From a_testbed.adv_emp
Group by salary WITH ROLLUP;
```

```
+-----------+---------------+
| EmpSalary | COUNT(emp_id) |
+-----------+---------------+
| 20000     |             1 |
| 22000     |             1 |
| 24000     |             1 |
| 25000     |             4 |
| 28000     |             6 |
| 30900     |             1 |
| 32000     |             2 |
| 45000     |             5 |
| all       |            21 |
+-----------+---------------+
9 rows in set (0.00 sec)
```

But that made the salary values into strings so they left align. We can use functions to get the alignment we want. Generally we prefer to leave the formatting of the output - such as alignment- to the application level.

```
Select coalesce(lpad(salary,12,' '), 'All Salaries') as salaryCol
, count(emp_id)
From a_testbed.adv_emp
Group by salary WITH ROLLUP;
+--------------+---------------+
| EmpSalary    | COUNT(emp_id) |
+--------------+---------------+
|        20000 |             1 |
|        22000 |             1 |
|        24000 |             1 |
|        25000 |             4 |
|        28000 |             6 |
|        30900 |             1 |
|        32000 |             2 |
|        45000 |             5 |
| All Salaries |            21 |
+--------------+---------------+
9 rows in set (0.00 sec)
```

Now we decide that we really do not need that much detail- after all there is no that much difference between a salary of 98000 and 98005. We decide we would rather have these grouped by 10K salary levels.

Demo 06: Truncate the salary to the 10K level

```
Select salary, truncate(salary, -4),  emp_id
From a_testbed.adv_emp
;
+--------+-----------------------+--------+
| salary | TRUNCATE(salary, - 4) | emp_id |
+--------+-----------------------+--------+
|  45000 |                 40000 |    101 |
|  20000 |                 20000 |    102 |
|  28000 |                 20000 |    103 |
|  25000 |                 20000 |    104 |
|  25000 |                 20000 |    105 |
|  28000 |                 20000 |    106 |
|  45000 |                 40000 |    107 |
|  28000 |                 20000 |    108 |
|  32000 |                 30000 |    109 |
|  45000 |                 40000 |    110 |
|  45000 |                 40000 |    111 |
```

```
|  30900 |                 30000 |    112 |
|  45000 |                 40000 |    113 |
|  32000 |                 30000 |    114 |
|  24000 |                 20000 |    115 |
|  28000 |                 20000 |    116 |
|  28000 |                 20000 |    117 |
|  25000 |                 20000 |    118 |
|  25000 |                 20000 |    119 |
|  22000 |                 20000 |    120 |
|  28000 |                 20000 |    121 |
+--------+-----------------------+--------+
21 rows in set (0.00 sec)
```

Demo 07: Use that query with the truncated salary and the emp id as a subquery and do the rollup on that

```
Select salary_10K, count(emp_id) as NumbrEmp
From (
   Select truncate(salary, -4)as salary_10K, emp_id
   From a_testbed.adv_emp) tbl_trunc
Group by salary_10K with rollup;
+------------+----------+
| salary_10K | NumbrEmp |
+------------+----------+
|      20000 |       13 |
|      30000 |        3 |
|      40000 |        5 |
|       NULL |       21 |
+------------+----------+
4 rows in set (0.00 sec)
```

We want to improve the display in the first column- if we had employees with salary under 10K, thie value in the first column would be 0 - which looks like they are not paid anything. So we want to display that message instead of 0.

(You can insert such a row here and then delete it after running the demos.)

It can help to use another level of subquery which will be used for the formatting; that way we avoid the formatting getting in the way of the group and rollup.

Demo 08: A three level query; there is no change in the display-yet

```
Select salary_10K, NumbrEmp
From (
   Select salary_10K, count(emp_id) as NumbrEmp
   From (
      Select truncate(salary, -4) as salary_10K, emp_id
      From a_testbed.adv_emp) tbl_trunc
      Group by salary_10K with rollup) rolled;
```

Demo 09: I have three things that could be displayed in the first column- (1)the salary value, or(2) "under 10K' for the first row with a salary of 0 and (3) 'Total if the first column is null. Do that with a Case.

```
Select case
    when salary_10K is null then '    Total'
    when salary_10K = 0  then 'Under 10K'
    else lpad(format(salary_10K,0), 9, ' ') end as salary_10K , NumbrEmp
```

```
From (
   Select salary_10K, count(emp_id) as NumbrEmp
   From (
      Select truncate(salary, -4) as salary_10K, emp_id
      From a_testbed.adv_emp) tbl_trunc
      Group by salary_10K with rollup) rolled;
+------------+----------+
| salary_10K | NumbrEmp |
+------------+----------+
|    20,000  |       13 |
|    30,000  |        3 |
|    40,000  |        5 |
|     Total  |       21 |
+------------+----------+
4 rows in set (0.00 sec)
```

2. Multilevel Group by with Rollups

Demo 10: Having more than one group by with Rollup. We get a rollup by year hired as well as for the departments and the whole table. This would not have been practical with a Union query. I highlighted the total rows

```
Select dept_id, year_hired, sum(salary), count(salary), count(emp_id)
From a_testbed.adv_emp
group by dept_id, year_hired WITH ROLLUP;
+---------+------------+-------------+---------------+---------------+
| dept_id | year_hired | SUM(salary) | COUNT(salary) | COUNT(emp_id) |
+---------+------------+-------------+---------------+---------------+
|      10 |       1980 |       45000 |             1 |             1 |
|      10 |       NULL |       45000 |             1 |             1 |
|      20 |       1990 |       20000 |             1 |             1 |
|      20 |       2000 |       78000 |             3 |             3 |
|      20 |       2010 |       30900 |             1 |             1 |
|      20 |       2013 |       22000 |             1 |             1 |
|      20 |       NULL |      150900 |             6 |             6 |
|      30 |       1990 |       28000 |             1 |             1 |
|      30 |       2008 |       45000 |             1 |             1 |
|      30 |       2010 |       24000 |             1 |             1 |
|      30 |       2012 |       56000 |             2 |             2 |
|      30 |       NULL |      153000 |             5 |             5 |
|      45 |       1995 |       28000 |             1 |             1 |
|      45 |       2000 |       73000 |             2 |             2 |
|      45 |       2008 |       32000 |             1 |             1 |
|      45 |       2010 |      122000 |             3 |             3 |
|      45 |       2012 |       25000 |             1 |             1 |
|      45 |       2013 |       25000 |             1 |             1 |
|      45 |       NULL |      305000 |             9 |             9 |
|    NULL |       NULL |      653900 |            21 |            21 |
+---------+------------+-------------+---------------+---------------+
20 rows in set (0.00 sec)
```

These are the rows for department 45.

```
Select emp_id, emp_name, year_hired, salary
From a_testbed.adv_emp
Where dept_id = 45
```

```
Order by year_hired;
+--------+----------+------------+--------+
| emp_id | emp_name | year_hired | salary |
+--------+----------+------------+--------+
|    121 | Goedel   |       1995 |  28000 |<< 1 emp, salary total is  28000
|    107 | Maddy    |       2000 |  45000 |<< 2 emp, salary total is  73000
|    108 | Pascal   |       2000 |  28000 |
|    109 | Boole    |       2008 |  32000 |<< 1 emp, salary total is  32000
|    111 | Turing   |       2010 |  45000 |<< 3 emp, salary total is 122000
|    113 | Lovelace |       2010 |  45000 |
|    114 | Polya    |       2010 |  32000 |
|    118 | Neumann  |       2012 |  25000 |<< 1 emp, salary total is  25000
|    119 | Wilkes   |       2013 |  25000 |<< 1 emp, salary total is  25000
+--------+----------+------------+--------+
9 rows in set (0.00 sec)
```

Let's look at some more examples. This is from the order entry tables in the a_oe database. I am going to use only orders in February 2013 to reduce the output.

Demo 11: Without RollUp: Group by customer and order and get the amount due and number of items per order. We have 8 orders in that time span.

```
Select cust_id, ord_id
, sum( quantity_ordered * quoted_price) as AmntDue
, sum( quantity_ordered) as NumberItems
From a_oe.order_headers
Join a_oe.order_details using (ord_id)
Where ord_date between '2013-02-01' and '2013-02-28'
Group by cust_id, ord_id
;
+---------+--------+---------+-------------+
| cust_id | ord_id | AmntDue | NumberItems |
+---------+--------+---------+-------------+
|  403000 |    508 |  383.90 |          10 |
|  403000 |    509 | 1049.93 |           7 |
|  403000 |    511 |  405.94 |           2 |
|  403050 |    507 |  145.99 |           1 |
|  404950 |    510 |   74.25 |           3 |
|  409150 |    515 |   49.99 |           1 |
|  409150 |    518 |  759.43 |           6 |
|  409150 |    716 |   12.95 |           1 |
+---------+--------+---------+-------------+
8 rows in set (0.00 sec)
```

Adding RollUp we get a summary for each grouping components- so we have totals by customer and also a grand total. This gives us 4 customer total lines and one grand total line.

Demo 12: With RollUp; I filtered on the date to reduce the display

```
Select cust_id, ord_id
, sum( quantity_ordered * quoted_price) as AmntDue
, sum( quantity_ordered) as NumberItems
From a_oe.order_headers
Join a_oe.order_details using (ord_id)
Where ord_date between '2013-02-01' and '2013-02-28'
Group by cust_id, ord_id with rollup
;
```

```
+---------+--------+---------+-------------+
| cust_id | ord_id | AmntDue | NumberItems |
+---------+--------+---------+-------------+
|  403000 |    508 |  383.90 |          10 |
|  403000 |    509 | 1049.93 |           7 |
|  403000 |    511 |  405.94 |           2 |
|  403000 |   NULL | 1839.77 |          19 | << cust total
|  403050 |    507 |  145.99 |           1 |
|  403050 |   NULL |  145.99 |           1 | << cust total
|  404950 |    510 |   74.25 |           3 |
|  404950 |   NULL |   74.25 |           3 | << cust total
|  409150 |    515 |   49.99 |           1 |
|  409150 |    518 |  759.43 |           6 |
|  409150 |    716 |   12.95 |           1 |
|  409150 |   NULL |  822.37 |           8 | << cust total
|    NULL |   NULL | 2882.38 |          31 | << Grand total
+---------+--------+---------+-------------+
13 rows in set (0.00 sec)
```

The rows with Null for just the order id are the customer totals-the total for all orders for that customer. There is no order id that would make sense for these rows since this is an aggregate. The row with null for the customer id is the total for all customers.

You need to pay attention to the MySQL conventions for the group by clause. In demo 11, we would get the same results if we did Group by ord_id since for each order id there is only one customer id. But for demo 12, if you use group by ord_id with rollup and do not include the cust_id, then you are asking for a roll up only by order id and you do not get the customer totals. You do get the grand total line at the bottom of the result set but it will show a cust id value which can be confusing.

```
+---------+--------+---------+-------------+
| cust_id | ord_id | AmntDue | NumberItems |
+---------+--------+---------+-------------+
|  403050 |    507 |  145.99 |           1 |
|  403000 |    508 |  383.90 |          10 |
|  403000 |    509 | 1049.93 |           7 |
|  404950 |    510 |   74.25 |           3 |
|  403000 |    511 |  405.94 |           2 |
|  409150 |    515 |   49.99 |           1 |
|  409150 |    516 |   12.95 |           1 |
|  409150 |    518 |  759.43 |           6 |
|  401890 |    519 |  114.74 |           6 |
|  404900 |    520 |   77.70 |           6 |
|  404900 |   NULL | 3074.82 |          43 |
+---------+--------+---------+-------------+
11 rows in set (0.00 sec)
```

Try working with this to get the following result

```
+----------+----------+---------+-------------+
| Customer | order_id | AmntDue | NumberItems |
+----------+----------+---------+-------------+
| 403050   | 507      |  145.99 |           1 |
| 403000   | 508      |  383.90 |          10 |
| 403000   | 509      | 1049.93 |           7 |
| 404950   | 510      |   74.25 |           3 |
| 403000   | 511      |  405.94 |           2 |
| 409150   | 515      |   49.99 |           1 |
| 409150   | 516      |   12.95 |           1 |
```

```
| 409150   | 518      |  759.43 |           6 |
| 401890   | 519      |  114.74 |           6 |
| 404900   | 520      |   77.70 |           6 |
| Total    | Total    | 3074.82 |          43 |
+----------+----------+---------+-------------+
11 rows in set (0.00 sec)
```

If you are in the habit of using the MySQL Group by convention- take more care with these queries.

Demo 13: You can work with functions to adjust the output; this is often easier if you use another level of subquery,

```
Select coalesce(cust_id, 'Grand Total') as CustID
, case when cust_id is null then ' '
       else coalesce(ord_id, 'Cust Total')
       end as OrderID
, AmntDue
, NumberItems
From (
   Select cust_id, ord_id
   , sum( quantity_ordered * quoted_price) as AmntDue
   , sum( quantity_ordered) as NumberItems
   From a_oe.order_headers
   Join a_oe.order_details using (ord_id)
   Where ord_date between '2013-02-01' and '2013-02-28'
   Group by cust_id, ord_id with rollup
) agg_data;
+-------------+------------+---------+-------------+
| CustID      | OrderID    | AmntDue | NumberItems |
+-------------+------------+---------+-------------+
| 403000      | 508        |  383.90 |          10 |
| 403000      | 509        | 1049.93 |           7 |
| 403000      | 511        |  405.94 |           2 |
| 403000      | Cust Total | 1839.77 |          19 |
| 403050      | 507        |  145.99 |           1 |
| 403050      | Cust Total |  145.99 |           1 |
| 404950      | 510        |   74.25 |           3 |
| 404950      | Cust Total |   74.25 |           3 |
| 409150      | 515        |   49.99 |           1 |
| 409150      | 518        |  759.43 |           6 |
| 409150      | 716        |   12.95 |           1 |
| 409150      | Cust Total |  822.37 |           8 |
| Grand Total |            | 2882.38 |          31 |
+-------------+------------+---------+-------------+
13 rows in set (0.00 sec)
```

Demo 14: Changing the order of the grouping items. Take demo 12 and reverse the order of the attributes in the Group by clause. If you were doing a plain Group By clause this would not make any difference; with the roll up it does. Since the ord_id was a finer level of grouping, we do not get any customer id totals. These are order totals.

```
Select cust_id, ord_id
, sum( quantity_ordered * quoted_price) as AmntDue
, sum( quantity_ordered) as NumberItems
From a_oe.order_headers
Join a_oe.order_details using (ord_id)
```

```
Where ord_date between '2013-02-01' and '2013-02-28'
Group by ord_id, cust_id  with rollup;
+---------+--------+---------+-------------+
| cust_id | ord_id | AmntDue | NumberItems |
+---------+--------+---------+-------------+
|  403050 |    507 |  145.99 |           1 |
|    NULL |    507 |  145.99 |           1 |
|  403000 |    508 |  383.90 |          10 |
|    NULL |    508 |  383.90 |          10 |
|  403000 |    509 | 1049.93 |           7 |
|    NULL |    509 | 1049.93 |           7 |
|  404950 |    510 |   74.25 |           3 |
|    NULL |    510 |   74.25 |           3 |
|  403000 |    511 |  405.94 |           2 |
|    NULL |    511 |  405.94 |           2 |
|  409150 |    515 |   49.99 |           1 |
|    NULL |    515 |   49.99 |           1 |
|  409150 |    518 |  759.43 |           6 |
|    NULL |    518 |  759.43 |           6 |
|  409150 |    716 |   12.95 |           1 |
|    NULL |    716 |   12.95 |           1 |
|    NULL |   NULL | 2882.38 |          31 |
+---------+--------+---------+-------------+
17 rows in set (0.00 sec)
```

Demo 15: Here we have three levels of groupings and you can see the same pattern of nulls. We have a rollup by each month in the year and a rollup by year and a grand total

```
Select
  YEAR(ord_date) as Order_year
, MONTH(ord_date) as Order_month
, Ord_id
, SUM(quantity_ordered * quoted_price) As AmntDue
, SUM(quantity_ordered) As NumberItems
From a_oe.order_headers
Join a_oe.order_details using (ord_id)
Group by year(ord_date), month(ord_date), ord_id with rollup;
+------------+-------------+--------+----------+-------------+
| Order_year | Order_month | Ord_id | AmntDue  | NumberItems |
+------------+-------------+--------+----------+-------------+
|       2012 |           2 |    516 |    12.95 |           1 |
|       2012 |           2 |   NULL |    12.95 |           1 |<< month total
|       2012 |          10 |    105 |  1205.40 |          29 |
|       2012 |          10 |    106 |   255.95 |           1 |
|       2012 |          10 |    107 |    49.99 |           1 |
|       2012 |          10 |    108 |    22.50 |           1 |
|       2012 |          10 |    109 |   149.99 |           1 |
|       2012 |          10 |    110 |   299.98 |           2 |
|       2012 |          10 |    400 |   465.00 |          20 |
|       2012 |          10 |    401 |   158.85 |           6 |
|       2012 |          10 |    402 |   158.85 |           6 |
|       2012 |          10 |   NULL |  2766.51 |          67 |<< month total
. . .   rows skipped
|       2011 |
|       2012 |          12 |    124 |    14.99 |           1 |
|       2012 |          12 |    125 |    55.00 |           1 |
|       2012 |          12 |    126 |    49.99 |           1 |
|       2012 |          12 |    127 |   124.98 |           3 |
|       2012 |          12 |    128 |   511.90 |           2 |
```

```
|         2012 |           12 |     129 |    299.97 |             3 |
|         2012 |           12 |     130 |   1145.00 |             3 |
|         2012 |           12 |    NULL |   2201.83 |            14 |<< month total
|         2012 |         NULL |    NULL |  12006.63 |           185 |<< year total
. . .   rows skipped
|         2013 |            5 |     525 |     52.77 |             4 |
|         2013 |            5 |     527 |    440.47 |            53 |
|         2013 |            5 |     528 |   2629.00 |            33 |
|         2013 |            5 |     529 |    279.49 |            11 |
|         2013 |            5 |     535 |    525.00 |             2 |
|         2013 |            5 |     536 |    525.00 |             2 |
|         2013 |            5 |    NULL |   4451.73 |           105 |<< month total
|         2013 |            6 |     301 |    205.00 |             1 |
|         2013 |            6 |     302 |    469.95 |            11 |
|         2013 |            6 |     303 |    125.00 |             1 |
|         2013 |            6 |     306 |   1000.00 |             2 |
|         2013 |            6 |     307 |   4500.00 |            10 |
|         2013 |            6 |     312 |   9405.00 |            50 |
|         2013 |            6 |     313 |    125.00 |             1 |
|         2013 |            6 |     324 |    599.00 |            20 |
|         2013 |            6 |     378 |   4500.00 |            10 |
|         2013 |            6 |     390 |   1400.00 |             8 |
|         2013 |            6 |     395 |   2925.00 |            15 |
|         2013 |            6 |     540 |    150.24 |             4 |
|         2013 |            6 |    NULL |  25404.19 |           133 |<< month total
|         2013 |         NULL |    NULL |  40417.31 |           336 |<< year total
|         NULL |         NULL |    NULL |  52423.94 |           521 |<< Grand total
+--------------+--------------+---------+-----------+---------------+
80 rows in set (0.00 sec)
```

2.1. Some Issues with RollUp

- With MySQL, you cannot use an Order by clause if you use With Rollup
- You can do a Descending sort of the Group By attributes; the totals are still at the bottom of each group.

You may have trouble trying to show only the total lines by simply adding a Having clause testing for nulls. This has to do with the way that MySQL implements this clause and the time at which it generates and adds the "null" rows. If you try to get just the totals by using the having clause here, you get no rows returned. This is one of those technique issues that causes problems with understanding the syntax rules.

Demo 16:

```
Select year(ord_date), month(ord_date), ord_id
, sum( quantity_ordered * quoted_price) as amntdue
, sum( quantity_ordered) as NumberItems
From a_oe.order_headers
Join a_oe.order_details using (ord_id)
Group by year(ord_date), month(ord_date), ord_id with rollup
Having ord_id is null;
```

Manual: Because the NULL values in the super-aggregate rows are placed into the result set at such a late stage in query processing, you cannot test them as NULL values within the query itself. For example, you cannot add HAVING product IS NULL to the query to eliminate from the output all but the super-aggregate rows.

On the other hand, the NULL values do appear as NULL on the client side and can be tested as such using any MySQL client programming interface.