Tutorial – Advanced OLAP

In this tutorial, you can use the ROBCOR data warehousing in the `dw` account. The tables that are relevant to the ROBCOR data warehouse are: `dw.charter_fact`, `dw.time`, `dw.pilot`, and `dw.model`.
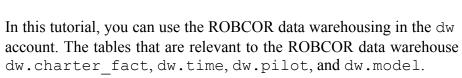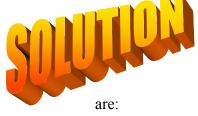
A. ROW_NUMBER(), DENSE_RANK(), RANK(), and PERCENT_RANK() Analysis Questions

1. The rank function displays the rank of a record. Its usage is as follow. To find the rank of the records in the time table is as follow:

   ```
   SELECT time_year, time_month,
      RANK() OVER (ORDER BY time_year, time_month) AS time_rank
   FROM dw.time;
   ```

2. Try the query below and compare the result with A.1. Investigate the purpose of using '+0' in order by time_month.

   ```
   SELECT time_year, time_month,
      RANK() OVER (ORDER BY time_year, time_month+0) AS time_rank
   FROM dw.time;
   ```

3. Display the row number of total charter hours used by each aircraft model in year 1996 (Hints: Use ROW_NUMBER() Over) The results should look like as follows.

   ```
   MOD_CODE    TIME_I SUM(TOT_CHAR_HOURS)     ROW_NUM
   ---------- ------ -------------------- ----------
   PA23-250   199608                  8.9          1
   C-90A      199602                 14.6          2
   PA23-250   199601                 16.9          3
   PA23-250   199605                 18.4          4
   C-90A      199603                 18.5          5
   PA23-250   199607                   19          6
   C-90A      199612                 19.2          7
   PA23-250   199604                 19.5          8
   PA23-250   199606                 19.5          9
   PA23-250   199602                 20.2         10
   PA23-250   199612                 20.7         11
   :
   :

   36 rows selected.
   ```

```
Select mod_code, time_id, sum(TOT_CHAR_HOURS),
    ROW_NUMBER() Over (Order By Sum(TOT_CHAR_HOURS)) AS Row_Num
From dw.Charter_Fact
Where time_id LIKE '1996%'
Group By mod_code, time_id;
```

4. Display the ranking of total charter hours used by each aircraft model in year 1996(Hints: Use `Dense_Rank() Over`) The results should look like as follows.

```
MOD_CODE    TIME_I SUM(TOT_CHAR_HOURS) DENSE_RANK
---------- ------ -------------------- ----------
PA23-250   199608                  8.9          1
C-90A      199602                 14.6          2
PA23-250   199601                 16.9          3
PA23-250   199605                 18.4          4
C-90A      199603                 18.5          5
PA23-250   199607                   19          6
C-90A      199612                 19.2          7
PA23-250   199604                 19.5          8
PA23-250   199606                 19.5          8
PA23-250   199602                 20.2          9
PA23-250   199612                 20.7         10
:
:
36 rows selected.
```

```
Select mod_code, time_id, sum(TOT_CHAR_HOURS),
    dense_rank() Over (Order By Sum(TOT_CHAR_HOURS)) AS dense_rank
From dw.Charter_Fact
Where time_id LIKE '1996%'
Group By mod_code, time_id;
```

5. Compare the result of A.3 and A.4, what is the difference?

6. Display the ranking of total charter hours used by each aircraft model in year 1996 (Hints: Use `Rank() Over`) The results should look like as follows.

```
MOD_CODE    TIME_I SUM(TOT_CHAR_HOURS)        RANK
---------- ------ -------------------- ----------
PA23-250   199608                  8.9          1
C-90A      199602                 14.6          2
PA23-250   199601                 16.9          3
PA23-250   199605                 18.4          4
```

```
C-90A        199603                         18.5              5
PA23-250     199607                           19              6
C-90A        199612                         19.2              7
PA23-250     199604                         19.5              8
PA23-250     199606                         19.5              8
PA23-250     199602                         20.2             10
PA23-250     199612                         20.7             11
:
:
36 rows selected.
```

```
Select mod_code, time_id, sum(TOT_CHAR_HOURS),
   Rank() Over (Order By Sum(TOT_CHAR_HOURS)) AS Rank
From dw.Charter_Fact
Where time_id LIKE '1996%'
Group By mod_code, time_id;
```

7. Compare the result of A.3, A.4 and A.6, what is the difference?

8. Modify the ranking in question A.6 above, where ranking based on Model, so that the results will look like this:

```
MOD_CODE    TIME_I SUM(TOT_CHAR_HOURS) RANK_BY_MODEL
---------- ------ -------------------- -------------
C-90A       199602               14.6              1
C-90A       199603               18.5              2
C-90A       199612               19.2              3
C-90A       199601               20.9              4
C-90A       199607               22.3              5
C-90A       199604                 23              6
C-90A       199608               24.1              7
C-90A       199605               24.1              7
C-90A       199609               24.8              9
C-90A       199610               24.9             10
C-90A       199611               26.3             11
...
36 rows selected.
```

```
Select mod_code, time_id, sum(TOT_CHAR_HOURS),
   Rank() Over (
      Partition By mod_code
      Order By Sum(TOT_CHAR_HOURS)) AS Rank_By_Model
From dw.Charter_Fact
Where time_id LIKE '1996%'
Group By mod_code, time_id;
```

9. Display the ranking of each airplane model based on the yearly total fuel-used and the ranking of yearly total fuel-used by each airplane model, and (Hints: use multiple partitioning ranking).

```
TIME MOD_CODE        TOTAL RANK_BY_YEAR RANK_BY_MODEL
---- ---------- ---------- ------------ -------------
1994 PA31-350   23773.8            1             3
1994 C-90A      16933.2            2             3
1994 PA23-250    9086.9            3             2
1995 PA31-350   24700.7            1             1
1995 C-90A      19058.8            2             1
1995 PA23-250    9133.9            3             1
1996 PA31-350   24052.3            1             2
1996 C-90A      17648.2            2             2
1996 PA23-250    8420.1            3             3
1997 PA31-350   11263.7            1             4
1997 C-90A       8068.2            2             4
1997 PA23-250    4624.1            3             4

12 rows selected.
```

```
SELECT t.time_year, f.mod_code,
       SUM(f.tot_fuel)as Total,
       RANK() OVER (PARTITION BY t.time_year
       ORDER BY SUM(f.tot_fuel)DESC) AS RANK_BY_YEAR,
       RANK() OVER (PARTITION BY f.mod_code
       ORDER BY SUM(f.tot_fuel)DESC) AS RANK_BY_MODEL
FROM dw.charter_fact f, dw.time t
WHERE f.time_id = t.time_id
GROUP BY t.time_year, f.mod_code;

TIME MOD_CODE        TOTAL RANK_BY_YEAR RANK_BY_MODEL
---- ---------- ---------- ------------ -------------
1994 PA31-350   23773.8            1             3
1994 C-90A      16933.2            2             3
1994 PA23-250    9086.9            3             2
1995 PA31-350   24700.7            1             1
1995 C-90A      19058.8            2             1
1995 PA23-250    9133.9            3             1
1996 PA31-350   24052.3            1             2
1996 C-90A      17648.2            2             2
1996 PA23-250    8420.1            3             3
1997 PA31-350   11263.7            1             4
1997 C-90A       8068.2            2             4
1997 PA23-250    4624.1            3             4

12 rows selected.
```

10. Using the rank function (nested within a sub query, because rank cannot exist in a where clause) display the mod_code and mod_name of the two airplanes that have the largest total fuel used.

```
MOD_CODE    MOD_NAME                      TOTAL     MYRANK
----------  --------------------  ----------  ----------
PA31-350    Navajo Chieftain        83790.5           1
C-90A       KingAir                 61708.4           2
```

```
SELECT *
FROM (
      SELECT m.mod_code, m.mod_name,
         SUM(f. tot_fuel) AS total,
          RANK() OVER (ORDER BY SUM(f. tot_fuel) DESC)
                                     AS myrank
      FROM dw.charter_fact f, dw.model m
      WHERE f.mod_code = m.mod_code
      GROUP BY m.mod_code, m.mod_name
)
WHERE myrank <=2;

MOD_CODE    MOD_NAME                      TOTAL     MYRANK
----------  --------------------  ----------  ----------
PA31-350    Navajo Chieftain        83790.5           1
C-90A       KingAir                 61708.4           2
```

11. Using the `Percent_Rank()` function (nested within a sub query), display the time periods which had revenue in the top 10% of the months.

```
TIME_I       TOTAL PERCENT_RANK
------  ----------  ------------
199503    51144.16             1
199408    49775.51    .975609756
199510    48538.01    .951219512
199409    47647.75    .926829268
199703    45872.32    .902439024
```

```
SELECT dw.time.time_id, Total, percent_rank
FROM (
  SELECT
    time_id,
    SUM(revenue) AS Total,
    PERCENT_RANK () OVER (ORDER BY SUM(revenue)) AS percent_rank
  FROM dw.charter_fact
  GROUP BY time_id
) t, dw.time
WHERE t.time_id = dw.time.time_id
AND percent_rank >= 0.9
ORDER BY percent_rank DESC;

TIME_I       TOTAL PERCENT_RANK
------  ----------  ------------
199503    51144.16             1
199408    49775.51    .975609756
199510    48538.01    .951219512
199409    47647.75    .926829268
199703    45872.32    .902439024
```

B. Cumulative and Moving Aggregate Questions

1. Use the cumulative aggregate to show the following results. We only need to show 1995 revenues (Hints: Since we only display 1995 data, there is no PARTITION).

```
TIME_I  SUM(REVENUE)  CUMMULATE_REV
------  ------------  -------------
199501      29575.47     29,575.47
199502      43279.39     72,854.86
199503      51144.16    123,999.02
...
12 rows selected
```

```
Select time_id, SUM(revenue),
   TO_CHAR(SUM(SUM(revenue))
   OVER(ORDER BY time_id ROWS UNBOUNDED PRECEDING),
   '9,999,999.99') AS Cummulative_Rev
From dw.Charter_Fact
Where time_id LIKE '1995%'
Group By time_id;
```

2. Redo question C.1 above, instead of using cumulative aggregate, use moving aggregate to show the following results moving aggregate of 3 monthly. (Hints: Use ROWS 2 PRECEDING).

```
TIME_I  SUM(REVENUE)  MOVING_3_MONT
------  ------------  -------------
199501      29575.47     29,575.47
199502      43279.39     36,427.43
199503      51144.16     41,333.01
199504      42129.05     45,517.53
199505      41943.54     45,072.25
199506      34230.05     39,434.21
...
12 rows selected
```

```
Select time_id, SUM(revenue),
   TO_CHAR(AVG(SUM(revenue))
   OVER(ORDER BY time_id ROWS 2 PRECEDING),
   '9,999,999.99') AS Moving_3_Months_Avg
From dw.Charter_Fact
Where time_id LIKE '1995%'
Group By time_id;
```

3. Compare the result of C.1 and C.2, investigate the difference between cumulative aggregate and moving aggregate.

4. Display the cumulative total fuel used based on the year, and another cumulative total used for each airplane model.

```
TIME MOD_CODE          TOTAL CUM_FUEL_YEAR CUM_FUEL_MODE
---- ---------- ---------- ------------- -------------
1994 C-90A        16933.2     16,933.20     16,933.20
1994 PA23-250      9086.9     26,020.10      9,086.90
1994 PA31-350     23773.8     49,793.90     23,773.80
1995 C-90A        19058.8     19,058.80     35,992.00
1995 PA23-250      9133.9     28,192.70     18,220.80
1995 PA31-350     24700.7     52,893.40     48,474.50
...
12 rows selected
```

```sql
SELECT t.time_year, f.mod_code,
       SUM(f.tot_fuel)as Total,
       TO_CHAR(SUM(SUM(f.tot_fuel))
          OVER(PARTITION BY t.time_year ORDER BY time_year
          ROWS UNBOUNDED PRECEDING),'9,999,999.99') AS
          Cum_fuel_year,
       TO_CHAR(SUM(SUM(f.tot_fuel))
          OVER(PARTITION BY f.mod_code ORDER BY f.mod_code
          ROWS UNBOUNDED PRECEDING),'9,999,999.99') AS
          Cum_fuel_model
FROM dw.charter_fact f, dw.time t
WHERE f.time_id = t.time_id
GROUP BY t.time_year, f.mod_code
ORDER BY time_year;

TIME MOD_CODE          TOTAL CUM_FUEL_YEAR CUM_FUEL_MODE
---- ---------- ---------- ------------- -------------
1994 C-90A        16933.2     16,933.20     16,933.20
1994 PA23-250      9086.9     26,020.10      9,086.90
1994 PA31-350     23773.8     49,793.90     23,773.80
1995 C-90A        19058.8     19,058.80     35,992.00
1995 PA23-250      9133.9     28,192.70     18,220.80
1995 PA31-350     24700.7     52,893.40     48,474.50
1996 C-90A        17648.2     17,648.20     53,640.20
1996 PA23-250      8420.1     26,068.30     26,640.90
1996 PA31-350     24052.3     50,120.60     72,526.80
:
:
12 rows selected.
```