OMSBA 5145

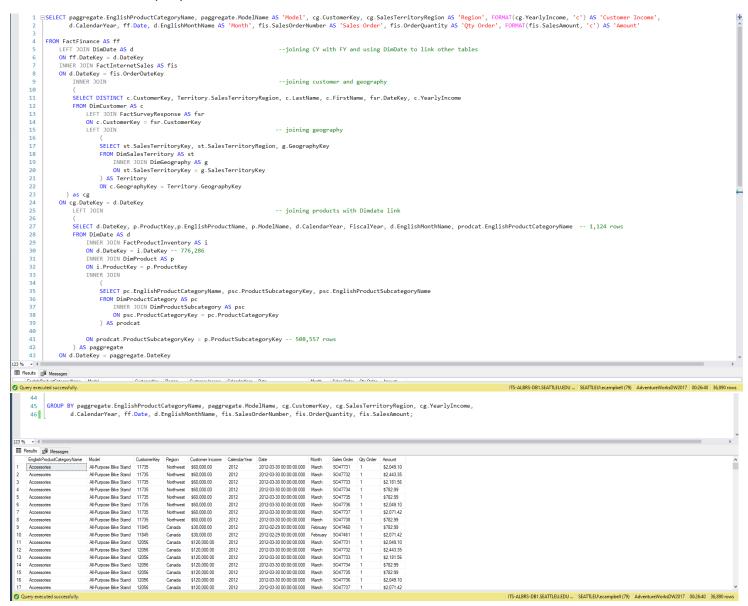
DATA TRANSLATION CHALLENGE

TECHNICAL PAPER ADVENTURE WORKS 2017

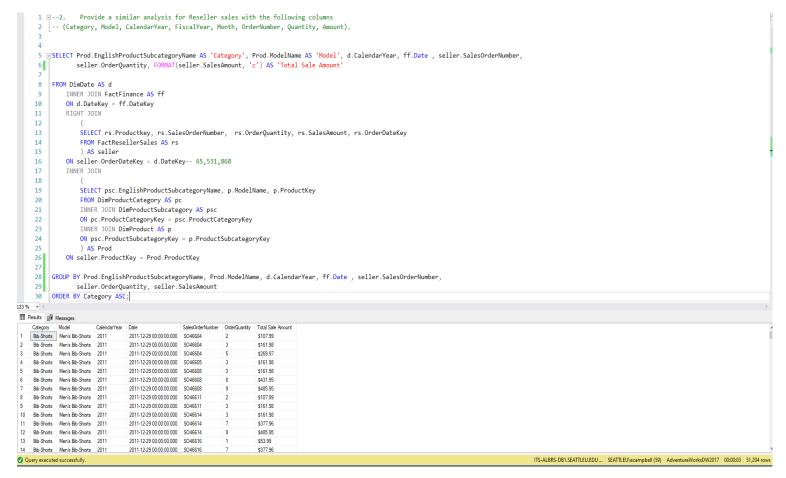
Ву

Campbell, Erik

The first business requirement was to evaluate internet sales and provide a query that would present the following the following columns: Category, Model, Customer Key, Region, Income, Calendar year, Fiscal year, Month, Sales Order Number, Sales Order Quantity and Sales Amount. There were several tables that needed to be joined and subqueries performed to get the variables requested. I determined that date was one variable common to all the tables and could be used to create linking tables. Thus, I decided to use FactFinance and DimDate in the FROM Clause and perform complex joins and subqueries from that root table in the FROM Clause. I joined the calendar year and fiscal year using date keys and then joined internet sales on a foreign date key. My next steps were joining this data with a subquery that grabbed data from tables containing geographical information, in turn could be used to link with customer data. This new data set then was left-joined with a subquery consisting of fields from three product-related tables. The global query resulted in 36,890 rows, but the query speed was suboptimal at 26 minutes to run during high server traffic times. The global query be further evaluated to examine if the sequence and/or order of subqueries in relation to the global query, or strategically placing more subqueries in the WHERE clause of the subqueries result in faster query response. I recommend that this query be run outside conventional business hours when there is less demand on the server. Or,



The Second business requirement was then to evaluate resellers while performing a similar analysis to internet sales; and ensure to include the following variables in the query results: Category, Model, Calendar Year, Fiscal Year, Month, Sales Orde Number, Sales Order Quantity, and Sales Order Amount. Once again, the Dimdate table and its date-key could prove useful in linking tables and I decided again to place the subqueries in the FROM Clause. The DimDate and Fact Finance tables were joined as a root linked table using the date-key and then joined with reseller table and applicable fields. Then, I inner-joined dates and reseller information with a sub-query selecting business requisite fields. Recognizing that business requirements required no customer or geographic data, I then grouped requisite business requirements query resulted in 51,204 rows in approximately 3 seconds. Analyst experimented with casting fiscal date values from Fact Finance table to year unsuccessfully, thus the whole fiscal date is present instead of just fiscal year.



Next business requirement led to the examination of total sales by year being rolled up by Territory Group and Country fields, but with UK being segregated from rollup and not being able to change data types. Inability to change data and wanting UK by itself had me thinking UNION. Thus, the aggregation of data was completed by using a union of internet sales and reseller sales that did not include UK data; and then doing the same for just UK data. I created queries for reseller sales and internet sales not excluding UK, and then just the UK. Then a union was used to aggregate UK Data from internet sales and reseller sales. A total of 39,379 rows resulted repeatedly in less than 10 seconds, but there are too many as the sales are not summed for a total using the Group By and Rollup by Sales Territory and Country. Analyst will have to ask supplementary questions about business requirement. A snapshot of code used is below to examine what went right and what analyst missing.

```
Show the total sales (overall) by year rolled up by the Territory group and country. A special request is that the United Kingdom
                             is no longer part of Europe and management wants to see their totals as a separate Territory group.
                             You cannot modify the data, so you will need to address this request in your query.
          6 SELECT FORMAT(SUM(fsr.SalesAmount), 'c') AS 'Total Sales', YEAR(fsr.OrderDate) AS 'Year', st.SalesTerritoryGroup, st.SalesTerritoryCountry
                  FROM FactResellerSales AS fsr
                       LEFT JOTN
                                SELECT g.CountryRegionCode, SalesTerritoryGroup, st.SalesTerritoryKey, st.SalesTerritoryCountry
         11
                       FROM DimGeography AS g
                              LEFT JOIN DimSalesTerritory AS st
        12
                               ON g.SalesTerritoryKey = st.SalesTerritoryKey
         14
                       WHERE st.SalesTerritorvKev <> '10'
         15
                                ) AS st
                       ON fsr.SalesTerritoryKey = st.SalesTerritoryKey
        17
                GROUP BY fsr.SalesAmount.fsr.OrderDate, fsr.SalesTerritorvKev.
                 ROLLUP(st.SalesTerritoryGroup, st.SalesTerritoryCountry)
         18
         19
                 UNTON
        20
        21
         22
                 SELECT FORMAT(SUM(fis.SalesAmount), 'c') AS 'Total Sales', YEAR(fis.OrderDate), cg.SalesTerritoryGroup, cg.SalesTerritoryCountry
                 FROM FactFinance AS ff
        23
                        LEFT JOIN DimDate AS d
                                                                                                                                          --joining CY with FY and using DimDate to link other tables
                        ON ff.DateKey = d.DateKey
                        INNER JOIN FactInternetSales AS fis
         26
        27
                        ON d.DateKey = fis.OrderDateKey
                            INNER JOIN
                                                                                                                                         --joining customer and geography
      29
                             SELECT DISTINCT c.CustomerKey, Territory.SalesTerritoryGroup, Territory.SalesTerritoryCountry, c.LastName, c.FirstName, fsr.DateKey, c.YearlyIncome, TerritoryCountry, c.LastName, c.FirstName, fsr.DateKey, c.YearlyIncome, fsr.DateKey, c.YearlyIncome, c.FirstName, c.Fir
      31
                            FROM DimCustomer AS c
                                     LEFT JOIN FactSurveyResponse AS fsr
      33
                                    ON c.CustomerKey = fsr.CustomerKey
                                    LEFT JOIN
                                                                                                                                        -- joining geography
      35
36
37
                                            SELECT st.SalesTerritoryKey, st.SalesTerritoryGroup, st.SalesTerritoryCountry, g.GeographyKey
                                           FROM DimSalesTerritory AS st
                                                   INNER JOIN DimGeography AS g
      39
40
                                                   ON st.SalesTerritoryKey = g.SalesTerritoryKey
                                            ) AS Territory
      41
                                           ON c.GeographyKey = Territory.GeographyKey
                          ) as cg
      42
               ON cg.DateKey = d.DateKey
WHERE cg.SalesTerritoryKey <> '10'
GROUP BY fis.SalesAmount, fis.OrderDate, cg.SalesTerritoryKey,
      43
      44
      45
46
               ROLLUP(cg.SalesTerritoryGroup, cg.SalesTerritoryCountry)
      47
48
      50
      51
               SELECT FORMAT(SUM(fis.SalesAmount), 'c') AS 'Total Sales', YEAR(fis.OrderDate), cg.SalesTerritoryGroup, cg.SalesTerritoryCountry
               FROM FactFinance AS ff
      52
53
                     M FACTFINANCE AS TT
LEFT JOIN DimDate AS d
ON ff.DateKey = d.DateKey
INNER JOIN FactInternetSales AS fis
                                                                                                                                         --joining CY with FY and using DimDate to link other tables
      54
      55
      56
57
                     --joining customer and geography
      58
59
                              SELECT DISTINCT c.CustomerKey, Territory.SalesTerritoryGroup, Territory.SalesTerritoryCountry, c.LastName, c.FirstName, fsr.DateKey, c.YearlyIncome, Territory.SalesTe
                             FROM DimCustomer AS c
                                    LEFT JOIN FactSurvevResponse AS fsr
      61
      62
                                     ON c.CustomerKey = fsr.CustomerKey
      63
                                    LEFT JOIN
                                                                                                                                        -- joining geography
       64
                                            SELECT st.SalesTerritoryKey, st.SalesTerritoryGroup, st.SalesTerritoryCountry, st.SalesTerritoryRegion, g.GeographyKey
FROM DimSalesTerritory AS st
INNER JOIN DimGeography AS g
ON st.SalesTerritoryKey = g.SalesTerritoryKey
       65
        66
67
68
       69
                                               AS Territory
                                            ON c.GeographyKey = Territory.GeographyKey
                ON c.GeographyKey = Territory.GeographyKey
ON cg.DateKey = d.DateKey
WHERE cg.SalesTerritoryKey = '10'
GROUP BY fis.SalesAmount, fis.OrderDate, cg.SalesTerritoryKey,
ROLLUP(cg.SalesTerritoryGroup, cg.SalesTerritoryCountry)
       71
72
73
74
75
76
77
78
79
                                                                                                                                                                                                                          - RESELLER UK ONLY-- 243,895 rows
                SELECT FORMAT(SUM(fsr.SalesAmount), 'c') AS 'Total Sales', YEAR(fsr.OrderDate), st.SalesTerritoryGroup, st.SalesTerritoryCountry
                 FROM FactResellerSales AS fsr
                               、
SELECT g.CountryRegionCode, st.SalesTerritoryGroup, st.SalesTerritoryCountry, st.SalesTerritoryKey
        83
84
                      SELECT g.CountryRegionCode, st.SalesTerritory(
FROM DimGeography AS g

LEFT JOIN DimSalesTerritory AS st

ON g.SalesTerritoryKey = st.SalesTerritoryKey

WHERE st.SalesTerritoryKey = '10'

) AS st
        85
86
87
        88
                       ON fsr.SalesTerritorvKev = st.SalesTerritorvKev
                GROUP BY Fsn.SalesTerritoryGroup, st.SalesTerritoryKey,
ROLLUP(st.SalesTerritoryGroup, st.SalesTerritoryCountry)
ORDER BY SalesTerritoryGroup DESC, SalesTerritoryCountry DESC;
123 %
       sults 🖼 Me
     Total Sales Year
$1,029,918.24 2013
                                Pacific
Pacific
                                                      Australia
Australia
       $1,144,353.60 2013
                                Pacific
                                                       Australia
       $1,175.52
                        2012
                                Pacific
                                                       Australia
       $1,199.76
                        2013
                                Pacific
                                                       Australia
       $1,263.36
                        2013
                                Pacific
                                                       Australia
       $1,295.76
                        2013
                                Pacific
       $1,319.76
                        2013
                                Pacific
       $1,335.60
                        2013
                                Pacific
      $1,437.12
                        2013
                                                                                                                                                                                                            ITS-ALBRS-DB1.SEATTLEU.EDU ... | SEATTLEU\ecampbell (60) | AdventureWorksDW2017 | 00:00:05 | 39,379 rd
```

The next business requirement was to evaluate internet sales and sales done by resellers by promotion and product. I wanted to look at promotional type in relation to sales made by our resellers and internet. Tables used were Promotion, internet sales and reseller sales and no complex joins completed. Emphasis of EDA was evaluating both internet sales and reseller sales under normal operations without a discount and with a promotion or discount. The query produced a result with 4 rows and gave us a view of sales with no discount those with promotions, but took a bit of time to run at 25 minutes. As stated earlier, performing a subquery or performing query outside of normal business hours might be prudent. Further EDA might would be to add fields from the Sales Reason and Internet sales Reason, products, customers, and geography tables and evaluate for seasonal and geographic trends in relation to promotion, but inventory forecasting, and management associated with trends.



The final business opportunity was evaluating the firm's customer demographics. The customer table(s) were joined with tables relating to geography and general summary statistics was performed. Summary statistics provide a glimpse into who target market and provide further research opportunities to explore methods of effectively reaching repeat customers with discretionary income. Thus, a query focusing on identifying customer clusters by geography, education, their income, and occupation was developed.

