University of Trieste
Data Management for Big Data Course
Academic Year 2022–2023

Data Warehouse case study

Davide Capone[*]   Sandro Junior Della Rovere[†]   Enrico Stefanel[‡]

Last update on 2023–06–21T12:28:33Z.

1   Introduction

1.1   TPC-H benckmark database

1.1.1   Database statistics

Lorem ipsum. . .

2   Export/import revenue value

Lorem ipsum. . .

3   Late delivery

It is asked to retrieve the number of orders where at least one "lineitem" has been received later than the committed date. The aggregation should

[*]davide.capone@studenti.units.it
[†]sandrojunior.dellarovere@studenti.units.it
[‡]enrico.stefanel@studenti.units.it

be performed with the Month → Year roll-up, and the (Customer's) Nation → Region roll-up.

Lorem ipsum...

4 Returned item loss

It is asked to retrieve the *revenue loss* for customers who might be having problems with the parts that are shipped to them, where a *revenue lost* is defined as `SUM(l_extendedprice*(1-l_discount))` for all qualifying "lineitems".

4.1 Naïve implementation

```
 1  WITH lineitem_orders AS (
 2      SELECT
 3          o_orderkey,
 4          l_partkey,
 5          l_suppkey,
 6          o_orderdate,
 7          o_custkey,
 8          l_extendedprice,
 9          l_discount,
10          l_returnflag,
11          l_commitdate,
12          l_receiptdate
13      FROM lineitem JOIN orders ON (l_orderkey = o_orderkey)
14  ),
15  query3 AS (
16  SELECT
17      EXTRACT (YEAR FROM o_orderdate) AS _year,
18      EXTRACT (QUARTER FROM o_orderdate) AS _quarter,
19      EXTRACT (MONTH FROM o_orderdate) AS _month,
20      c_name,
21      SUM(l_extendedprice * (1 - l_discount)) AS returnloss
22  FROM
23      lineitem_orders
24      JOIN customer ON o_custkey = c_custkey
25  WHERE
26      l_returnflag = 'R'
```

```
27        -- AND c_name = 'Customer#000129976'
28        -- AND EXTRACT (QUARTER FROM o_orderdate) = 1
29  GROUP BY
30      _year,
31      _quarter,
32      _month,
33      c_custkey,
34      c_name
35  )
36  SELECT * FROM query3;
```