20/15

ADAM BRONIEWSKI
00540635

Université libre de Bruxelles
Faculté des Sciences Appliquées

January 25, 2022

INFO-H-415 – Advanced databases

First session examination

Spatial Databases

The Research and Innovative Technology Administration (RITA) coordinates the U.S. Department of Transportation's (DOT) research programs. It collects several statistics about many kinds of transportation means, including the information about flight segments between airports summarized by month. It is possible to download a set of CSV files in ZIP format, one by year, ranging from 1990 up until now. These files include information about the scheduled and actually departured flights, the number of seats sold, the freight transported, and the distance traveled, among other ones. Figure 1 shows a relational data warehouse built with this data.

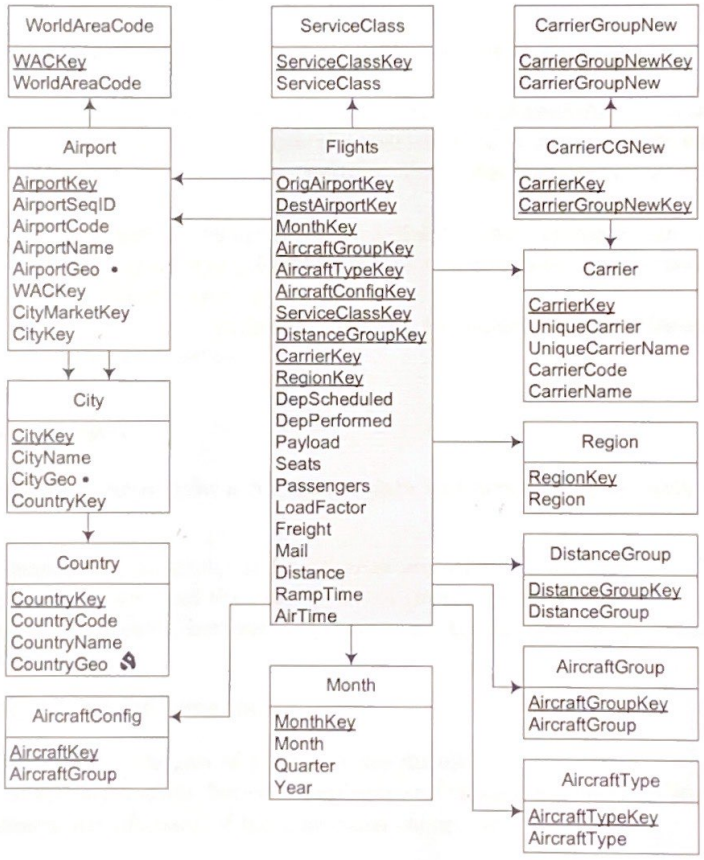

Figure 1: A relational schema for the RITA data warehouse

00540635

Write in SQL the following queries.

1. Display the spatial union of all airports with more than 5,000 departures in 2012.
2. Display for each city the three closest airports and their distance to the city independently of the country in which the city and the airport are located.
   N.B. A possible solution for this query uses the window functions in SQL.
3. Give the total number of persons arriving to or departing from airports closer than 15 km from the city center in 2012.
4. Give for 2012 the ratio between the number of persons arriving to or departing from aiports closer than 15 km from the city center and the number of persons arriving to or departing from airports located between 15 and 40 km from the city center.
5. For cities operated by more than one airport, give the total number of arriving and departing passengers at the airport closest to the city center, and the ratio between this value and the city total.

You will find below some of the common functions in PostGIS.

- float ST_Area(geometry): Returns the area of the surface of geometry.
- boolean ST_Within(geometry, geometry): Returns true if geometry A is completely inside geometry B.
- boolean ST_Contains(geometry, geometry): Returns true if geometry B is completely inside geometry A.
- boolean ST_Intersects(geometry, geometry): Returns true if geometry B intersects geometry A.
- geometry ST_Union(geometry): Returns the spatial union from a set of geometries.
- geometry ST_GeomFromText(string): Returns a specified geometry value from Well-Known Text representation (WKT).
- string ST_AsText(geometry): Returns the Well-Known Text representation of the geometry.
- geometry ST_Centroid(geometry): Returns the geometric center of a geometry, or equivalently, the center of mass of the geometry as a POINT.
- ~~integer~~ ST_Distance(geometry, geometry): Returns the minimum 2D Cartesian distance between two geometries in projected units.

Temporal Databases

Figure 2 shows a relational schema for a Northwind data warehouse extended with temporal features. In particular

- The lifespan employees, products, and categories are kept.
- The unit price of products and the description of categories are temporal.
- The supervision relationship between employees and the assignment of employees to cities are temporal relationships.

Enforce with triggers the following constraints.

1. All periods defining the lifespan of a product are disjoint.
2. The lifespan of the relationship between products and categories (ProductCategory) is covered by the ~~intersection of~~ the lifespans of the concerned categories.

Write in SQL the following queries.

3. For each employee, total sales amount of products she sold with unit price greater than $30 at the time of the sale.
4. For each product, list the name, unit price, and total sales amount by month.
5. Total sales amount for employees assigned to only one city. Notice that the assignment of employees to cities is represented in the Territories table.

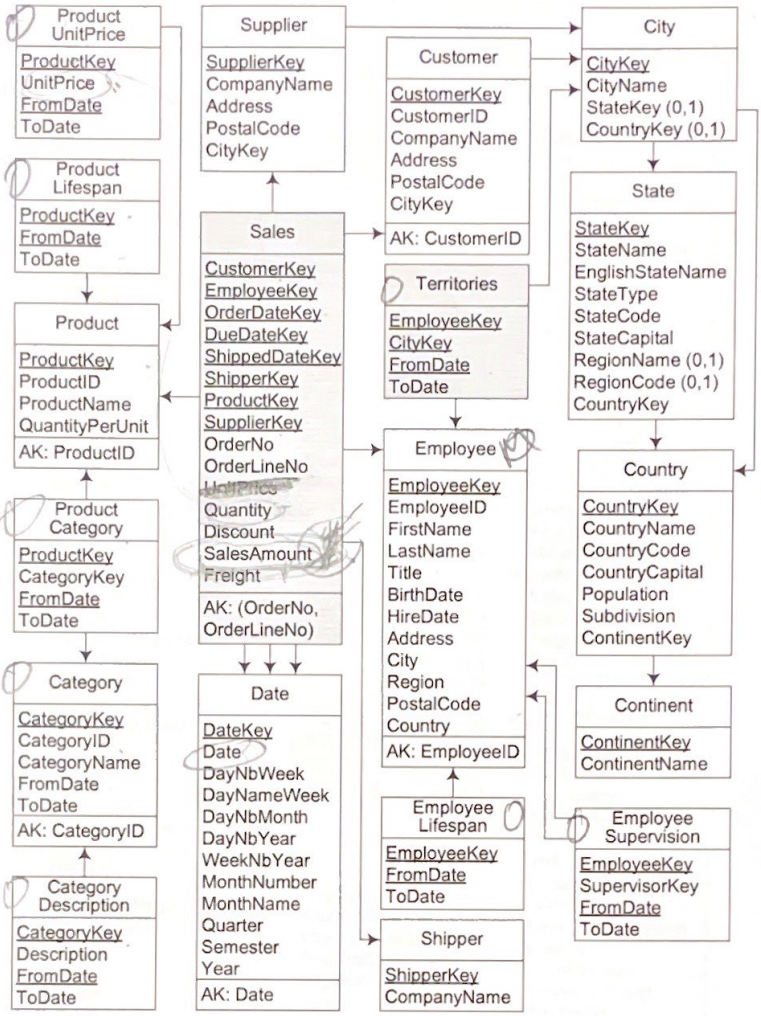

Figure 2: A relational schema for a temporal extension of the Northwind data warehouse

Graph Databases

Consider the relational schema of the Foodmart data warehouse given in Fig. 3 and its implementation in Neo4j given in Fig. 4. Write in Cypher the following queries.

1. All measures for stores in the states of California and Washington summarized at the state level.
2. All measures for the top-five store cities based on sales count.
3. Unit sales and number of customers by product subcategory.
4. Unit sales by customer city and percentage of the unit sales of the city with respect to its state.
5. Sales profit in 2017 by store type and store city, for cities whose unit sales in 2017 exceeded 25,000.

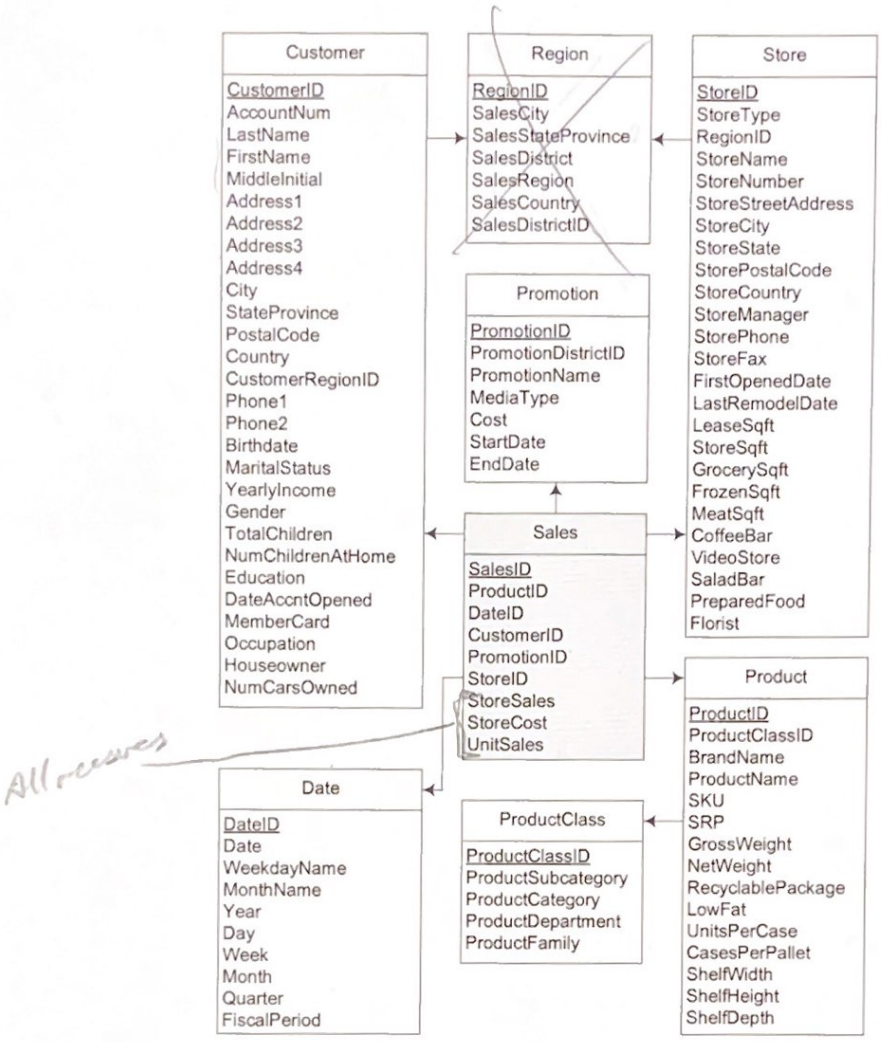

Figure 3: Schema of the Foodmart relational data warehouse

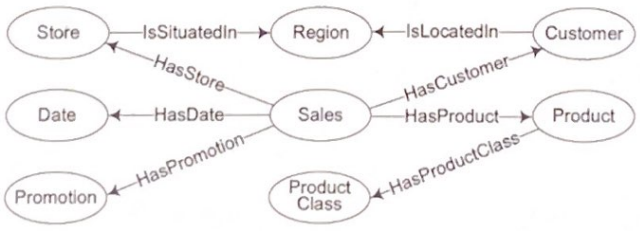

Figure 4: Schema of the Foodmart data warehouse in Neo4j