

Exercises Data Warehousing Multidimensional Databases

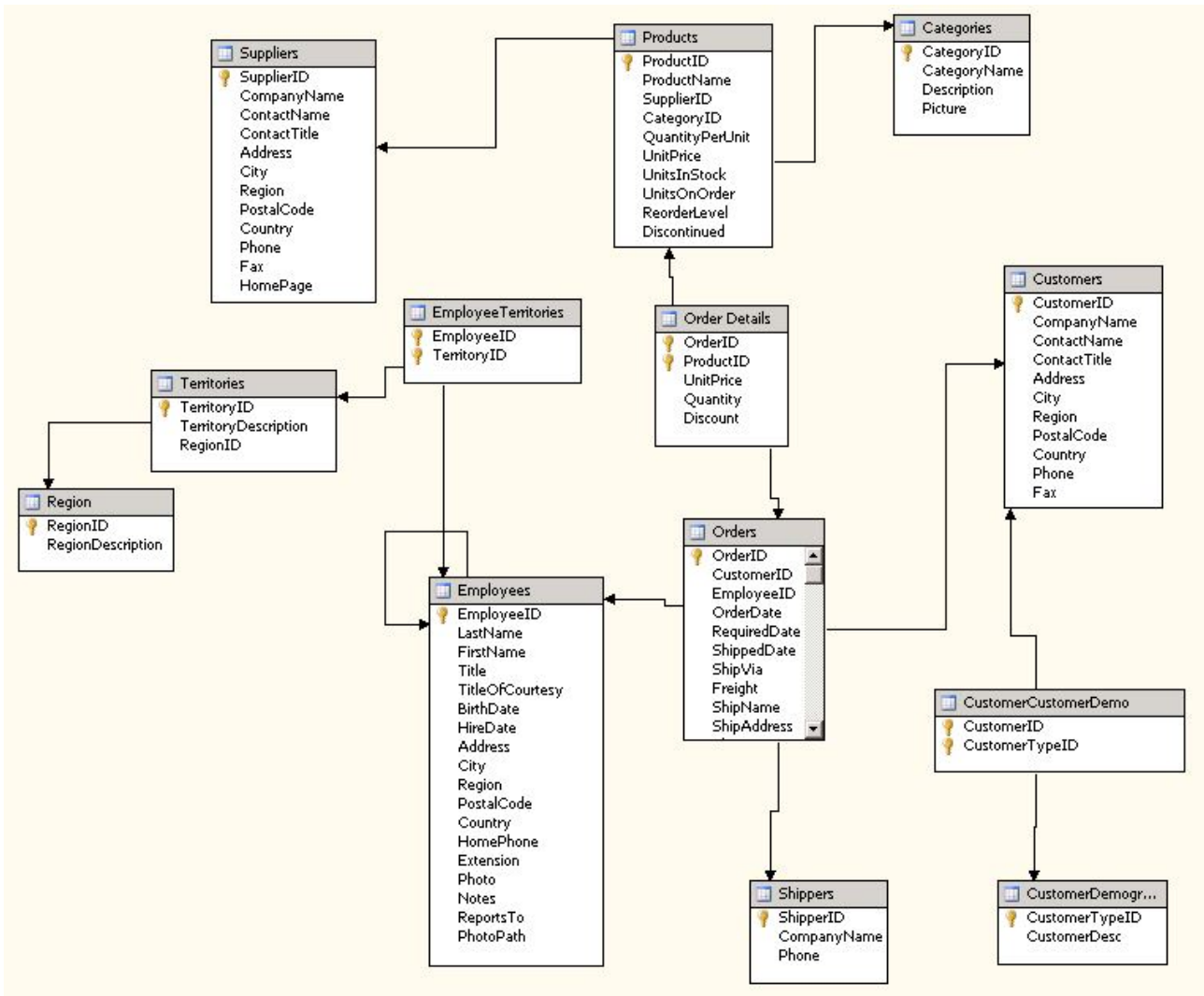


Figure 1: Schema of the Northwind database

1. Consider the Northwind database whose schema is given in Figure 1. This database contains information of orders placed by customers. For every order the detail is given of what products were sold, for what unit price and in what quantity. The employee that secured the order is recorded as well as the date in which the order was inserted. For customers the city they live in etc. is recorded, and for employees their salesdistrict. For this database, create queries to generate the following reports:
 - (a) Select the number of sales per category and country.
 - (b) Select the 3 top-selling categories overall (hint: use “select top 3” construction).
 - (c) Produce an overview of sales by month for these categories (hint: get month and year with “month” and “year” functions). Are there countries and product categories for which the trend over time is increasing?
 - (d) List total amount of sales in \$ by employee and year (discount in OrderDetails is at UnitPrice level). Which employees have an increase in sales over the three reported years?
 - (e) Get an individual sales report by month for employee 9 (Dodsworth) in 1997.
 - (f) Get a sales report by country and month.
2. The sales department of a supermarket chain wants to have a system to support the strategic planning and evaluation of promotions. To this end, they need sales information over the different stores of the supermarket chain. For their analysis tasks they want to compute average sales and total sales, for different

products, either at product level or brand level, for different stores at different levels of granularity: individual store, province where the store is located, and country, and for different time periods: per year, month, quarter, semester and also by day of the week.

- (a) How would you conceptually model the data needed by the sales department as a data cube? E.g., what are the measures, the dimensional attributes, the hierarchies, the aggregations that are needed?
- (b) Given the cube of (a), explain how you would construct the answers to the following queries with the operations slice-and-dice, pivot, roll-up, and drill-down. If necessary, indicate in which cell(s) of the constructed cube the answer can be found:
 - i. Give the total overall sales per store.
 - ii. Give an overview of the average sales per month per province.
 - iii. Give the subcube with only dimensions store at level province and day at level month for the average and total sales for the period 1999 till 2005.
- (c) Give an SQL:1999 expression that produces the datacube (i.e., contains all aggregates of the cube using the null value in an attribute to represent aggregation on the corresponding dimension). How do you handle the multiple measures? The hierarchy?

3. Give SQL:1999 expressions for the queries in 2(b).

4. Suppose that we have a relation Sales(Product, Month, Store, Amount). There are five products: P1, P2, P3, P4, P5, 12 months of data, and three stores: S1, S2, and S3.

- (a) (Dense setting) Suppose that every product has been sold in every month in every store; i.e., for every combination of a product p , a month m , and a store s , there is a tuple (p, m, s, a) with a non-zero amount.
 - i. How many tuples does this relation contain?
 - ii. How many tuples does a data cube with dimensions Product, Month, Store, and measure Amount contain?
- (b) (Sparse setting) Consider the following (sparse) relation:

Product	Month	Store	Amount
P1	Jan	S1	a1
P1	Jan	S2	a2
P2	Feb	S2	a3
P2	Feb	S3	a4
P3	Jan	S1	a5
P3	Feb	S1	a6
P4	Feb	S1	a7
P5	Jan	S3	a8

How many non-empty cells does the data cube of this relation contain?