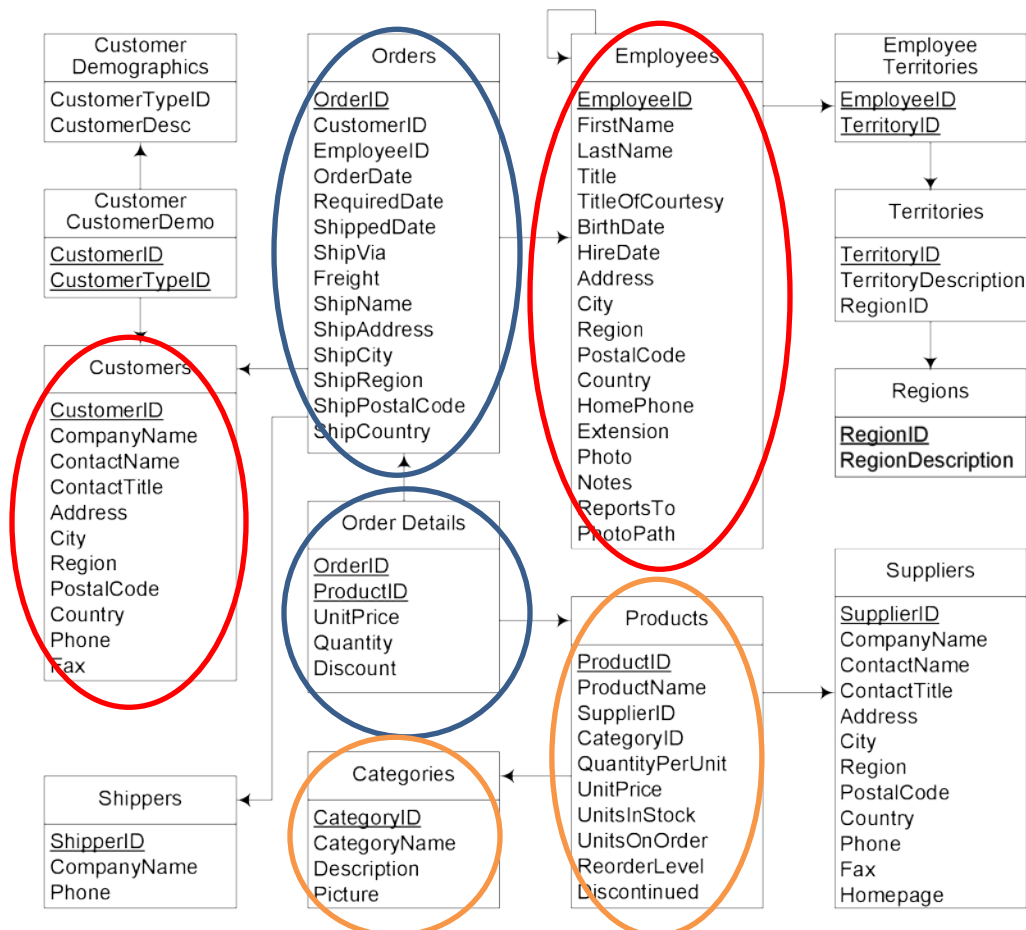


## SSIS Exercise – ETL workflow to update the database

1. Now write a package that updates the dimension and fact table. Keep in mind that changes in the dimensions Customer and Employee are treated with the type-2 strategy.
  - a. Suppose that we are tracking the changes by file comparison. In this way we can generate for every table in the source database, three sets of tuples:
    - i. tuples that were added
    - ii. tuples that were updated
    - iii. tuples that were deleted
  - b. Your package can assume the existence of these difference tables. This is the starting point of the exercise. The database containing some changes is available on the database server under the name Northwind-updates. In the following diagram of the operational Northwind database, the tables for which the differences are monitored have been indicated. For the blue tables we only consider inserts. For the orange tables there can be inserts and updates, but they have to be treated as type-1 changes (error corrections). For the red tables there are inserts, deletes, and type-2 updates.



Define now, on the basis of the differences tables, a package that updates the current data warehouse. For simplicity you may assume that:

- a. *Every* change in dimensions employee and customer is a type-2 change;
- b. *Every* change in the product dimension is a type-1 change;
- c. The date dimension is *never* updated. If needed, this table is updated manually by the database administrator.
- d. *Disregard the recursive hierarchy* on the employee dimension. Consider the correct treatment of this recursive hierarchy in the view of the type-2 changes as an extension if you finish your exercises early.

For bulk updates it is advisable to gather all tuples to be updated in a temporary relation and use an SQL statement to update the complete relation at once. **Be careful not to introduce duplicate tuples into the data warehouse during the updating process.**