DWH Assignment

A.A. 2024/2025

- Davide Ticchiarelli
- Niccolò Ciotti
- Giampaolo Marino
- Luca Renzi
- Riccardo Angelini

Preliminary Overview

Database: The database contains sales data from Northwind Traders, a fictitious company specializing in food import-export.

Goal: Analysis of orders based on time, employees, customer (also geographical area) and product.

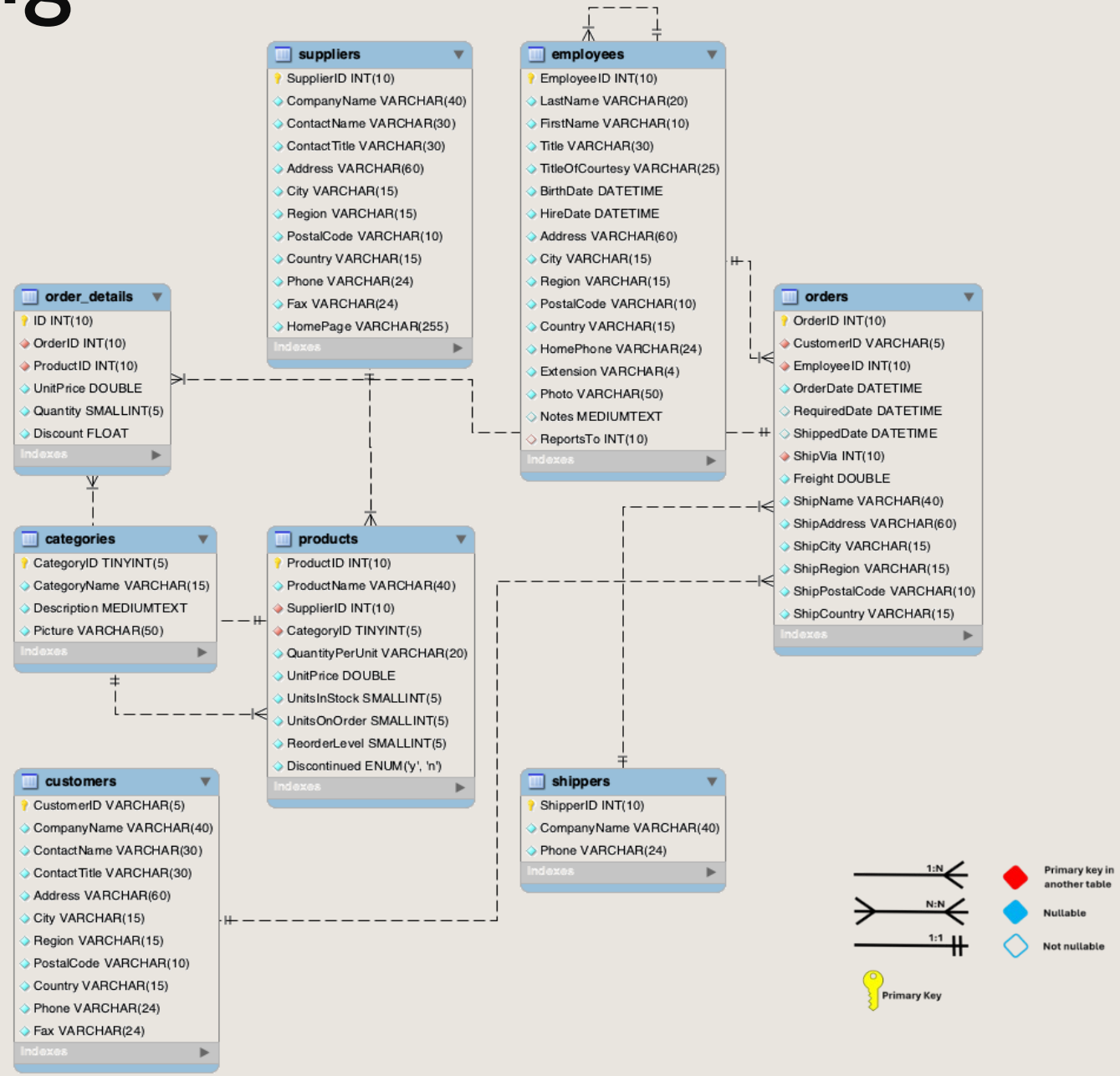
Measures:

1. Calculation of average revenue
2. Calculation of total number of orders
3. Determination of products sold by order
4. Calculation of average delivery time

Reverse engineering

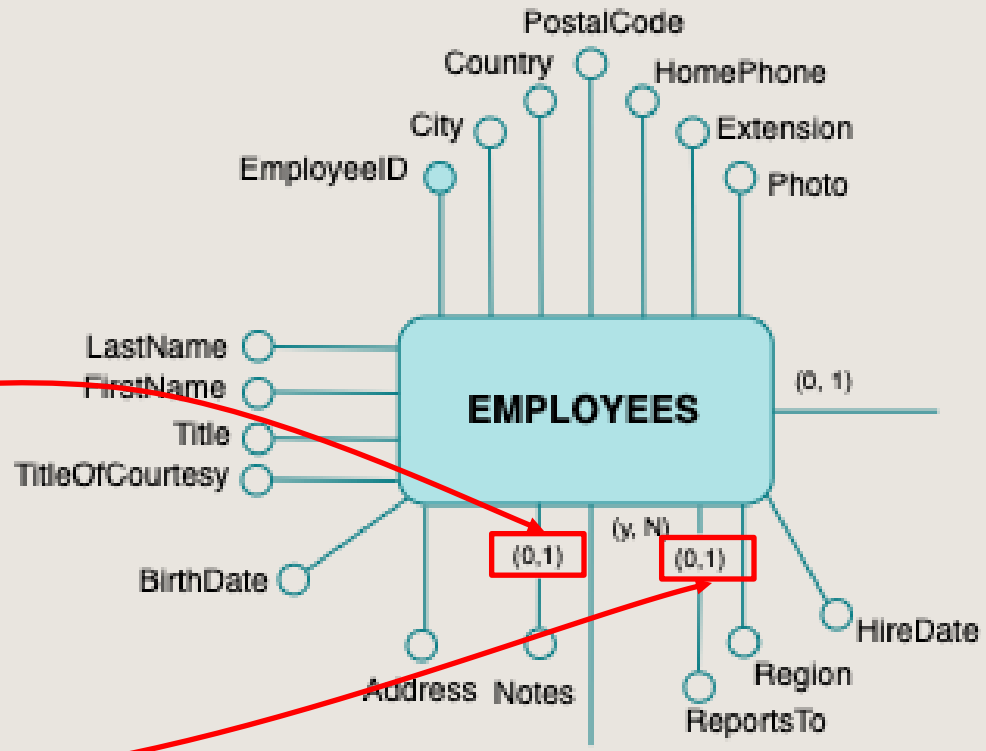
The ER model was generated through the following steps:

- Analysis of the SQL file and generation of the ER schema by reverse engineering
- Identification of entities and relationships
- Primary and foreign key recognition
- Construction of the E-R model

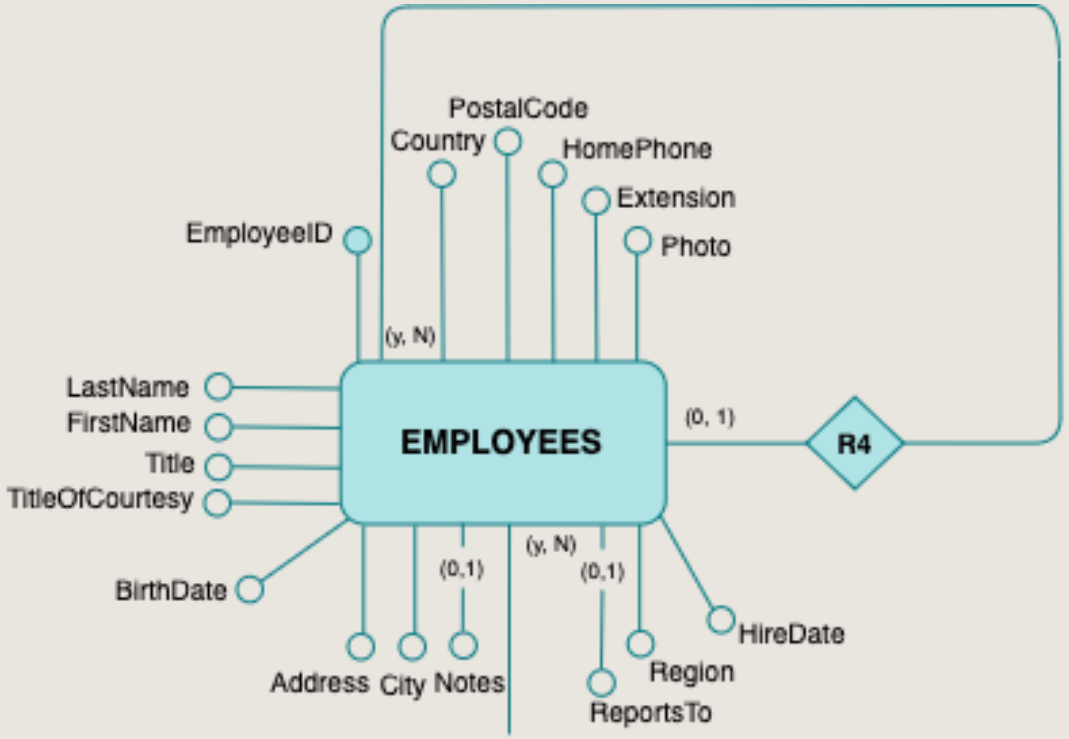


Optional attributes- Example

```
CREATE TABLE `employees` (
  `EmployeeID` int(10) unsigned NOT NULL AUTO_INCREMENT,
  `LastName` varchar(20) CHARACTER SET utf8 COLLATE utf8_unicode_ci NOT NULL DEFAULT '',
  `FirstName` varchar(10) CHARACTER SET utf8 COLLATE utf8_unicode_ci NOT NULL DEFAULT '',
  `Title` varchar(30) CHARACTER SET utf8 COLLATE utf8_unicode_ci NOT NULL DEFAULT '',
  `TitleOfCourtesy` varchar(25) CHARACTER SET utf8 COLLATE utf8_unicode_ci NOT NULL DEFAULT '',
  `BirthDate` datetime NOT NULL,
  `HireDate` datetime NOT NULL,
  `Address` varchar(60) CHARACTER SET utf8 COLLATE utf8_unicode_ci NOT NULL DEFAULT '',
  `City` varchar(15) CHARACTER SET utf8 COLLATE utf8_unicode_ci NOT NULL DEFAULT '',
  `Region` varchar(15) CHARACTER SET utf8 COLLATE utf8_unicode_ci NOT NULL DEFAULT '',
  `PostalCode` varchar(10) CHARACTER SET utf8 COLLATE utf8_unicode_ci NOT NULL DEFAULT '',
  `Country` varchar(15) CHARACTER SET utf8 COLLATE utf8_unicode_ci NOT NULL DEFAULT '',
  `HomePhone` varchar(24) CHARACTER SET utf8 COLLATE utf8_unicode_ci NOT NULL DEFAULT '',
  `Extension` varchar(4) CHARACTER SET utf8 COLLATE utf8_unicode_ci NOT NULL DEFAULT '',
  `Photo` varchar(50) CHARACTER SET utf8 COLLATE utf8_unicode_ci NOT NULL DEFAULT '',
  `Notes` mediumtext CHARACTER SET utf8 COLLATE utf8_unicode_ci,
  `ReportsTo` int(10) unsigned DEFAULT NULL,
  PRIMARY KEY (`EmployeeID`),
  KEY `idx_employees_lastname` (`LastName`),
  KEY `idx_employees_postalcode` (`PostalCode`),
  KEY `idx_ReportsTo` (`ReportsTo`),
  CONSTRAINT `FK_employees_reports_to` FOREIGN KEY (`ReportsTo`) REFERENCES `employees` (`EmployeeID`)
)
```



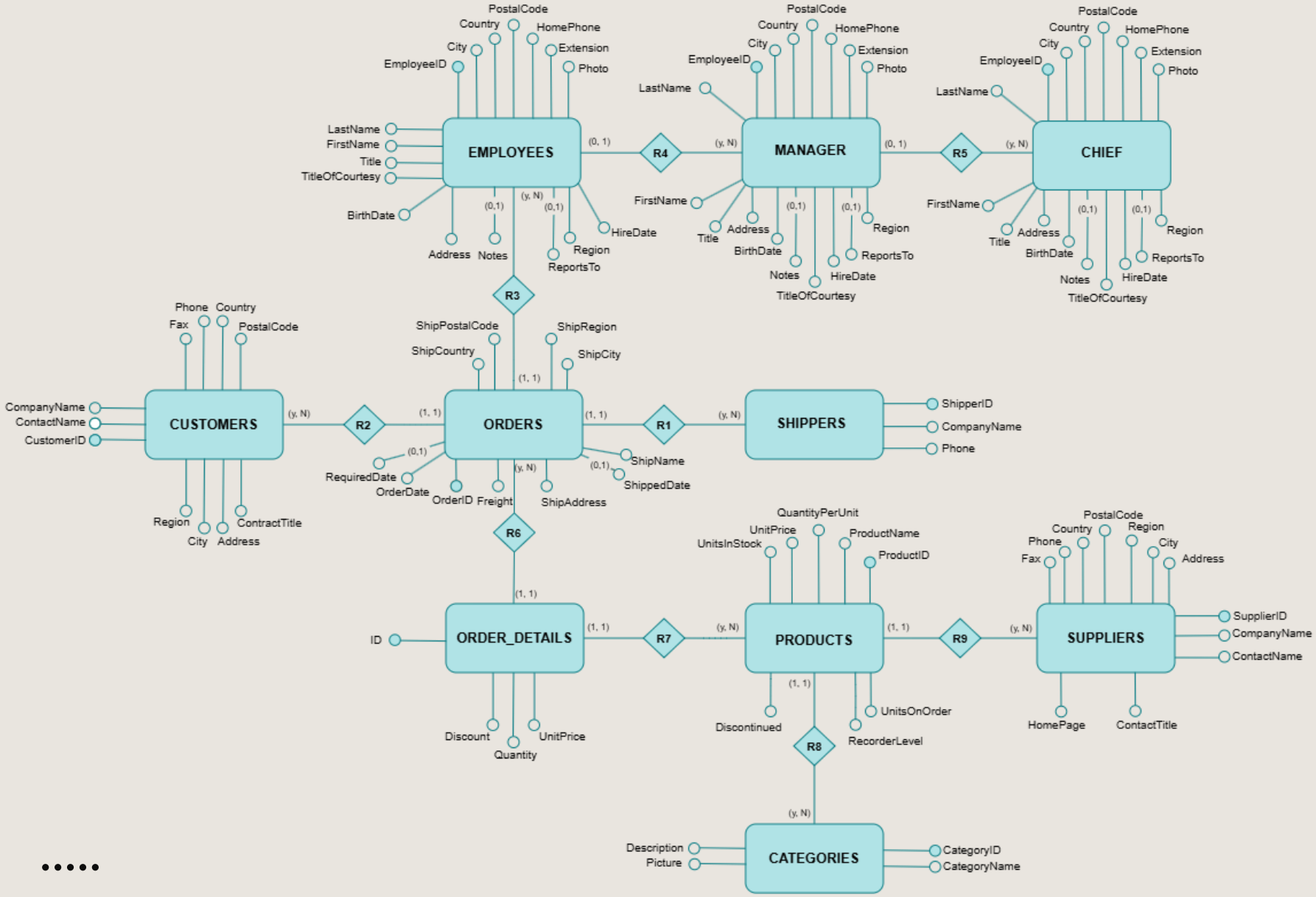
Cyclical relationship



ReportsTo	Title
2	Sales Representative
NULL	Vice President, Sales
2	Sales Representative
2	Sales Representative
2	Sales Manager
5	Sales Representative
5	Sales Representative
2	Inside Sales Coordinator
5	Sales Representative

The **R4** relationship has been transformed into a specific sequence of relationships, explaining their different levels. In particular, the following entities were created based on the above result: **EMPLOYEES**, **MANAGER** and **CHIEF**

Final ER model

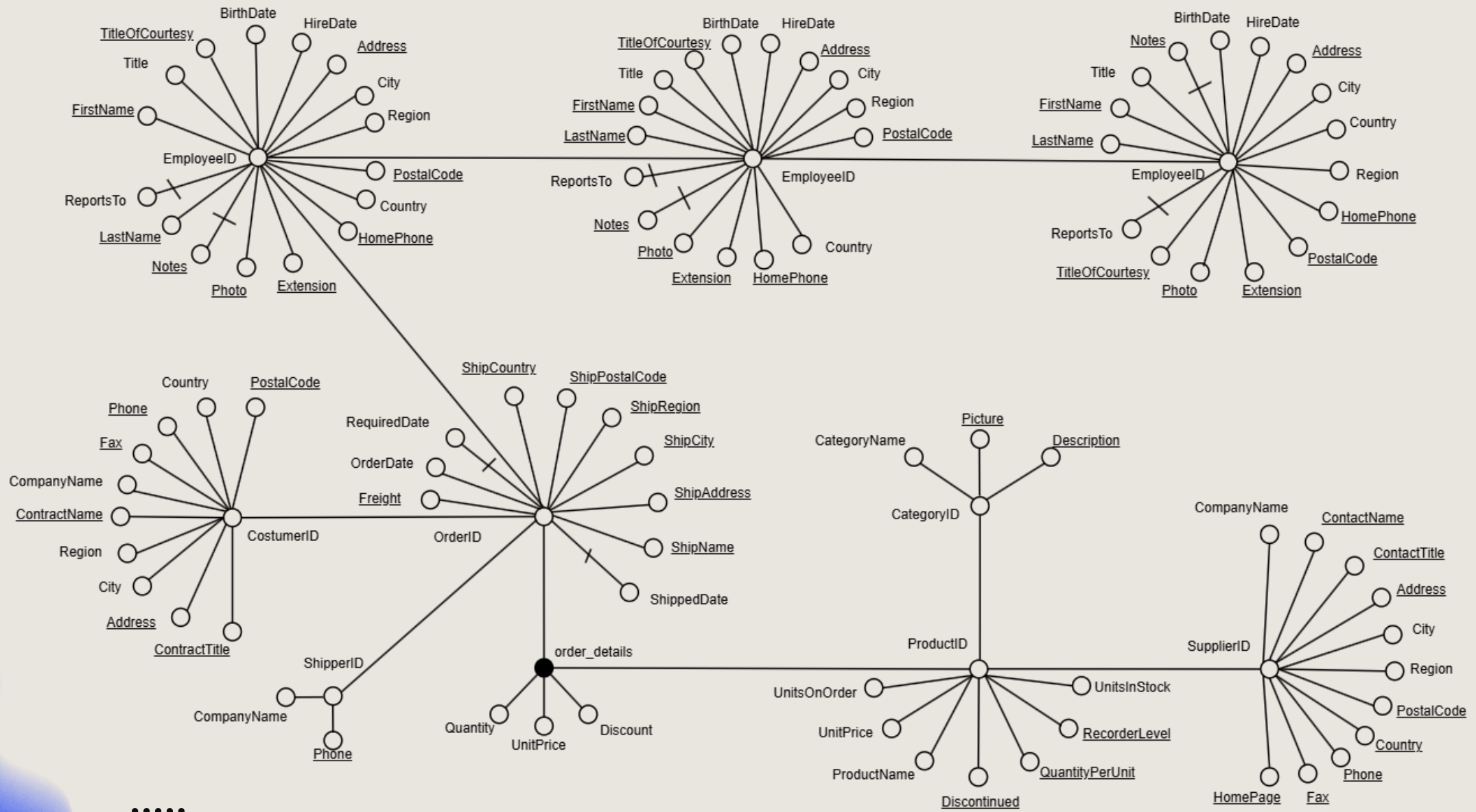


Attributes tree

Starting from the ER model, the tree was generated by the following procedure:

- **Identification of the fact:** creation of the root node of the tree from its primary key. In our case, we used "**ORDER_DETAILS**" as a fact of the attribute tree to analyze orders
- **Definition of attributes**
- **Iterative tree extension**

Attributes tree

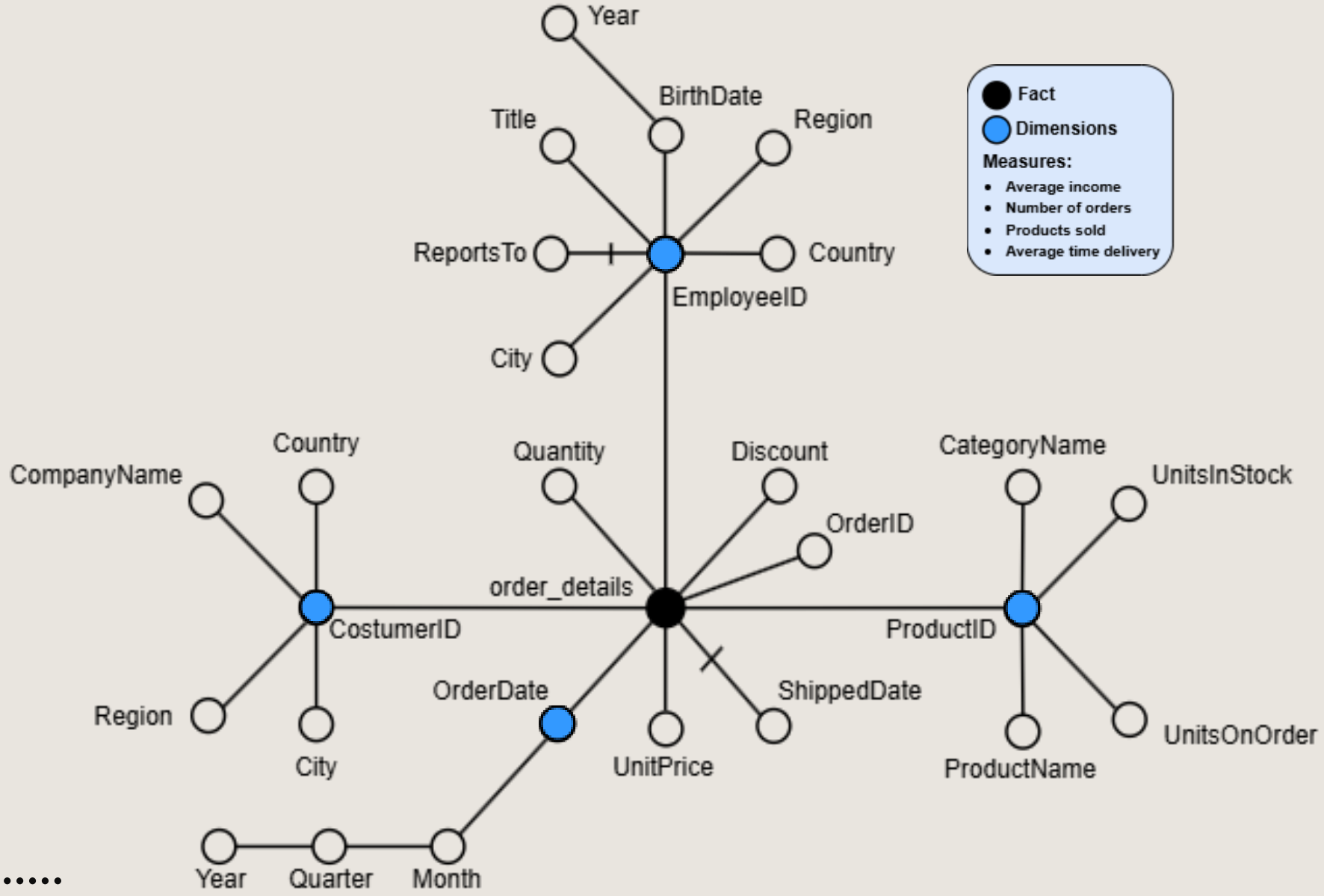


Refinement of the tree

In the following list we have reported all the operations performed to reduce the attribute tree:

- All elements considered irrelevant for the analysis have been eliminated. For example, nodes related to **MANAGER**, **CHIEF**, **SHIPPERS** and **SUPPLIERS**, as well as some attributes of **PRODUCTS**, **CUSTOMERS**, **CATEGORIES**, **ORDERS** and **EMPLOYEES**
- The **CategoryName** node has been **grafted** to **CategoryID**
- Functional dependencies have been **eliminated**:
  - **OrderID -> CustomersID**
  - **OrderID -> EmployeesID**
  - **OrderID -> OrderDate**
  - **OrderID -> ShippedDate**
- The **BirthDate** and **OrderDate** nodes have been **reified** to clarify their hierarchy

Refined attributes tree



Measures analysis

- Average income: **avg**(**sum**(order_details.UnitPrice * order_details.Quantity * (1 - order_details.Discount)))
- Number of orders: **count**(order_details.OrderID)
- Products sold by order: **sum**(order_details.Quantity)
- Average delivery time : **avg**(**datediff**(order_details.ShippedDate, order_details.OrderDate))

	Product		Customer		Employee		Time	
	Add.	**Aggr. Op.**	**Add.**	**Aggr. Op.**	**Add.**	**Aggr. Op.**	**Add.**	**Aggr. Op.**
Average income	No	Algebraic	No	Algebraic	No	Algebraic	No	Algebraic
Number of orders	No	Distributive	No	Distributive	No	Distributive	No	Distributive
Products sold by orders	Yes	Distributive	Yes	Distributive	Yes	Distributive	Yes	Distributive
Average time delivery	No	Algebraic	No	Algebraic	No	Algebraic	No	Algebraic

Thanks for your attention