Introduction

In this project, you will be working with the Northwind database. The Northwind database was originally created by Microsoft for demonstration purposes. The database contains the sales data for *Northwind Traders*, a fictitious import/export company that sells specialty foods.

In this project, you are asked to perform two main tasks:
- Create an integrated view of the employee data that exists in the Northwind database and in the "company" database that you used in the labs.
- Create a data warehouse from the Northwind database, and write an interesting query that can be used for analysis and reporting purposes.

The next sections describe in more detail what you should do in each of these tasks.

The Northwind database

The script **northwind.sql** contains the SQL instructions needed to create the Northwind database. The database schema is shown in the figure below, where arrows represent foreign keys between tables.
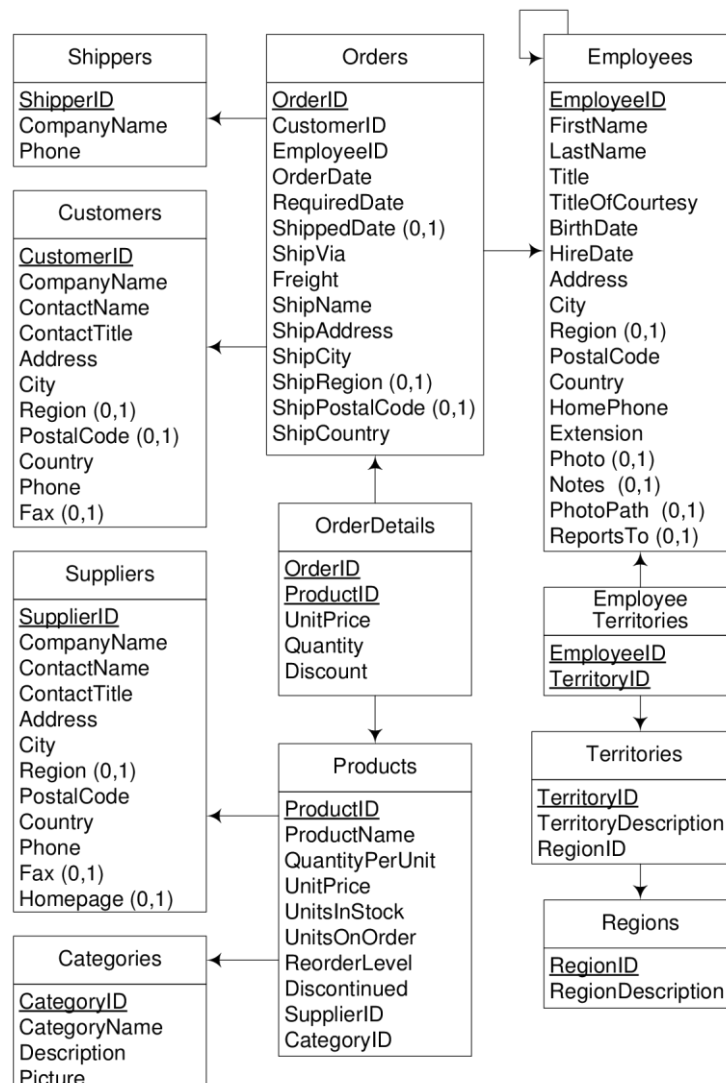
In this project, we are focusing on the information about employees, orders, customers, products, and suppliers. Note that each order belongs to a single customer, but it may contain multiple products. Products are organized into different categories, and each product is provided by a certain supplier. Customers and suppliers have the same sort of geographical data associated with them, such as city, country, etc.

Integrating the employee data

In the labs, you have worked with a "company" database that contained information about branches, departments and employees. Perhaps not all of that information can be mapped to the Northwind database. However, some of that information can. Your goal is to find the maximum extent of information that can be mapped between the two databases.

Specifically, in this task you should present:
1. The schema matching between both databases. List the correspondences between each table/column in the Northwind database and each table/column in the "company" database.
2. The common (mediated) schema between both databases.
3. The schema mapping (i.e. views) to retrieve data from both databases at the same time.
4. A transformation to detect approximate duplicates between the job titles in both databases. The output of this transformation should be a list of pairs of potential duplicates.

Creating the data warehouse

In the labs, you have created a data warehouse from the steelwheels database. This included developing an ETL process in Pentaho Data Integration (PDI), defining the OLAP cube in Pentaho Schema Workbench (PSW), querying the data warehouse in Saiku Analytics, and creating some reports in Pentaho Report Designer (PRD). In this project, you should follow a similar approach to create a data warehouse from the Northwind database.

The data warehouse should have a star schema with the following dimensions:
- Customer with company name, city and country.
- Product with product name and category name (slowly-changing dimension).
- Supplier with company name, city and country.
- Time with day, month and year.

Use sales and quantity as measures. Sales is calculated as UnitPrice*Quantity*(1-Discount).

In this task, you should present:
5. The SQL instructions needed to create the data warehouse tables.
6. The transformations (and a job) to implement the ETL process in PDI.
7. The XML cube definition created with PSW.
8. An interesting MDX query and a screenshot of its results as displayed in Saiku. The query should not be too simple nor too complicated. It should demonstrate some of the capabilities of the MDX language. As a suggestion, use WHERE or WITH and at least one MDX function.
9. A report created with PRD. The report can be based on the same or on a different MDX query. As a suggestion, the report could contain a list and a chart.

Preparing to submit the project

After you complete the tasks above, use a word processor (such as LibreOffice Writer or similar) to prepare a document with the following sections:

- At the top of the first page, write your name(s) and student number(s).
1. Present the schema mapping in the form of a table or diagram.
2. Present the mediated schema in text format.
3. Present the SQL views in text format, but formatted and indented in a way that makes it easy to read.
4. Take a screenshot of your transformation. Also, take a screenshot of the configuration window for each step. Finally, take a screenshot of the Preview window for the output step. You can use Alt+PrintScreen to capture these windows.
5. Present the SQL instructions in text format, but formatted and indented in a way that makes it easy to read.
6. Take a screenshot of each transformation/job, followed by a screenshot of the configuration window for each step. You can use Alt+PrintScreen to capture these windows.
7. Present the XML code in text format, but formatted and indented in a way that makes it easy to read.
8. Present the MDX query in text format, followed by a screenshot of the results.
9. Present the MDX query in text format (if different from the previous one). Take two screenshots of the report in PRD: one in Design mode and another in Preview mode.

Submitting the project

Save the document as a PDF file (without image compression, or with lossless compression) and submit it in Fénix until December 13, 2017.

Note

Use this project as part of your preparation for the exam. There is no point in submitting a project without fully understanding how it works. The more attention you dedicate to this project, the better you will be prepared. Good luck!