Business Intelligence Lab

Written test 10/6/2015

Deliver solutions within 4 h

Notice: use your own SQL Server credentials (the lbi account is disabled)

Exercise 1 (8 pts). Let $C(i)$ be the set of customers that bought product i. Develop a SSIS package reading `sales_fact` from the *foodmart* database, which outputs a CSV file with a single column containing the list of products i such that at least 10% of customers in $C(i)$ is from customers in $C(925)$.

What to deliver: SSDT solution.

Exercise 2 (8 pts). Solve Exercise 1 by developing a Java program Bought.java using JDBC. Only `SELECT` (without `WHERE`/`ORDER BY`/`GROUP BY` statements) from a single table are allowed in SQL queries.

What to deliver: Bought.java, myJDBCdef.props (with only the parameters needed for a test of the program).

Exercise 3 (8 pts). Write a single MDX query that solves Exercise 1.

What to deliver: MDX query and a brief comment about it, PowerPoint file with the screenshot of the query result.

Exercise 4 (8 pts). Generalize the problem from Exercise 1 to consider customers of a specific sex and/or country, e.g., $C_{sex=F}(i)$ is the set of female customers that bought product i. Solve the generalized problem using association rules. Experiment with Weka.

What to deliver: either a Weka knowledge flow `.kfml` file or a PowerPoint file with screenshots of Weka explorer or a Java program with Weka API calls, and a brief description of the steps of the analysis.

How to deliver: send an e-mail with a single <your surname>.zip file attached to ruggieri@di.unipi.it, including your name, surname, student ID, and computer IP address (http://www.whatismyip.com).

Results and oral exam. Results will be emailed to the students shortly, including the date and time for those who are admitted to the oral exam.