



UNIVERSITY  
OF SKÖVDE

School of Informatics (IIT)

## WRITTEN EXAMINATION

Course **Data Warehousing - teknologier och metoder**

Sub-course

Course code **IT382G**

Credits for written examination **6**

Date **2024-03-22**

Examination time **8:15 – 12:30**

Examination responsible **Manfred Jeusfeld**

Teachers concerned **Manfred Jeusfeld**

Aid at the exam/appendices

Students are allowed to bring a Swedish-English dictionary to the exam

- Instructions
- Take a new sheet of paper for each exam part.
  - Take a new sheet of paper when starting a new question.
  - Write only on one side of the paper.
  - Write your name and personal ID No. on all pages you hand in.
  - Use page numbering.
  - Don't use a red pen.
  - Mark answered questions with a cross on the cover sheet.

Grade points: 90 (30 for each of the three parts)

Answer in Swedish or English. Answer all the sub-questions of a question. To pass the exam you must reach at least 50% of the possible points. The ECTS grades are calculated as follows from the percentage of points you achieved (90 points=100%):

<50%: F; 50-59%: E; 60-69%: D; 70-79%: C; 80-89%: B; 90-100%: A

**Examination results shall be made available according to university rules via LADOK**

*Good luck!*

Total number of pages excluding this page: 3



UNIVERSITY  
OF SKÖVDE

### Part 1: Central concepts (30 points)

- a) (10) The standard data warehouse architecture(s) include the so-called **data staging area**. What is the purpose of this component? From where does it get its data, what is done with the data at the data staging area, and where does the data go to from the data staging area. Provide a detailed answer and include a diagram that shows the position of the data staging area in the data warehouse architecture.
- b) (10) So-called **information package** diagrams are used to collect the requirements for a data warehouse. Explain what is represented in such diagrams. Give an example that has at least three dimensions and one metric. Is it possible to have several metrics in the same information package diagram?
- c) (10) Low **data quality** in data warehouses leads to wrong results. Which types of errors in data can occur? Explain at least **three types of “data corruption”** and give examples for these types. How can errors be corrected? Which component of the data warehouse is responsible for correcting errors? How can we be sure that the data in a data warehouse is complete?



UNIVERSITY  
OF SKÖVDE

## Part 2: Data Warehousing and OLAP (30 points)

- a) (10 points) Create a two-dimensional **PIVOT** table for the dimensions time and product out of the following **fact table**. Note that the price attribute is the measurement attribute here.

Time	Location	Product	Customer	Price
10:31	Skövde	Cykel-A	Fred	10 500
12:30	Malmö	Cykel-A	Mary	15 800
12:47	Malmö	Bil-B	Fred	118000

- b) (10) Certain measurement attributes are not **summarizable**, i.e. the application of the SUM operation makes no sense. Explain what we understand by “**flow observations**”, “**stock observations**” and “**value-per-unit observations**”. Give an example of a measurement attribute that cannot be summarized. Define the term “**semi-summarizable**” and give an illustrative example.
- c) (10) Dimension tables need to be constructed in a way that the aggregation (roll-up) is done correctly, in particular for avoiding **double-counting or missing values** in the aggregation (SUM,AVG,...). Explain what has to be checked for dimension tables in order to avoid the two problems. Give an example of a malformed dimension tables where double-counting can occur.



UNIVERSITY  
OF SKÖVDE

### Part 3: Data Mining (30 points)

- a) (10) Discuss the **main differences of OLAP versus “data mining”**. Which technique is more automated? What types of knowledge can be found in data mining that cannot be easily found with OLAP (or multi-dimensional queries)? Which one is more suitable for predicting what happens in the future? List at least three main differences!
- b) (10 points) What are **neural networks**? Give an example diagram of a neural network. What is the role of the weights? What is the purpose of “training” the neural network? Answer all parts!
- c) (10) Explain the idea and principal procedure of the **“memory-based reasoning”** method? What is the goal of the method? Give an example to illustrate the method.