

School of Informatics (IIT)

WRITTEN EXAMINATION

Course **Data Warehousing - teknologier och metoder**

Sub-course

Course code **IT382G**

Credits for written examination **6**

Date **2026-06-08**

Examination time **8:15 – 12:30**

Examination responsible **Manfred Jeusfeld, Mikael Berndtsson**

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 questions. To pass the exam, all three parts must be passed. Each question has an equal weight of 10 points. The final grade is calculated from these points.

All **three** part solutions must have at least 50% of the possible points, i.e. 15 points for each part. 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 should be made public within 18 working days

Good luck!

Total number of pages excluding this page: 3

Part 1: Central concepts (30 points)

- a) (10 points) The relational schema of a data warehouse can follow the so-called “star schema” or the “snowflake schema” approach. Create a **star schema** as a UML class diagram for a company that sells products. Include at least three dimensions (time, customer, product, shop,...). How can a star schema be converted to a snowflake schema?
- b) (10 points) Dimension tables sometimes need to be updated, e.g. when a new customer is added. They may also need to be updated, e.g. when the address of a customer changes. The course discussed **three strategies to make such updates**. Describe these strategies and give an example on how they work in the case of updating the address of a customer. What are the advantages and disadvantages of the strategies!
- c) (10 points) **Information packages** are a tool to capture requirements to a data warehouse. Consider a logistics company that transports containers on trucks between harbors, warehouses and customers. Among others, the bus company is interested in the time it takes to transport certain container types from/to certain location types. Create an information package diagram (table) that has at least three dimensions and three attributes per dimension.

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	11000
12:30	Malmö	Cykel-B	Mary	16000
12:47	Skara	Bil-B	Paul	120000
13:23	Skövde	Bil-C	Anne	230000

- b) (10 points) 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.
- c) (10 points) **MOLAP** and **ROLAP** are two strategies to provide data cubes to data warehouse clients (such as OLAP tools). Explain what we understand by MOLAP and ROLAP! When should MOLAP be used and when is ROLAP a better choice?

Part 3: Data Mining (30 points)

- a) (10 points) 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 points) Explain the idea and principal procedure of the **“k-Means”** method? For what does the “k” stand for? What is a centroid in the context of k-Means?