

# Database Systems

A.Y. 2025/2026

Master's degree in computer science, third written exam 20/02/2026 - 9:30 to 14:00.

“ConferenceHub Inc.”	
1	Each conference organized by the company "ConferenceHub Inc." is identified by an acronym and has an associated name, the venue where it will take place, the URL of the homepage, and an optional set of sponsors who finance the conference. For each sponsor, the following information is known: name, date the funding was provided, and amount. Each conference is assigned a set of organizers who form the program committee. For each organizer, the following is known: name, affiliation, address, phone number, and email. Each organizer may indicate, for each conference in which they participate as a program committee member, one or more research areas denoted by an acronym and a description. Articles submitted to a conference (an article cannot be submitted to more than one conference) are characterized by a sequential number within the conference, a title, and one or more authors for whom the following information must be maintained: name, affiliation, address, phone number, and email. Note that among the authors, one must be designated as the "contact author" who will receive all communications regarding the submitted article. Articles are also divided into different categories: tutorial, short papers, posters, industrial papers, and research papers. For industrial papers, information (name and address) about all partners who contributed to the research presented in the article must be stored. Based on the specified research area, articles are assigned to program committee members (the number of reviewers per article varies from a minimum of two to a maximum of four) who must prepare a review. The review prepared by a reviewer for a particular article includes a score on originality, significance, and quality of the proposed work, as well as a global score and comments that will be sent to the contact author. Based on the collected evaluations, the status of each article may be either "accepted" or "rejected".

- 1) analyze these specifications, filtering the ambiguities present and then grouping them homogeneously.  
Represent the specifications with an ER diagram. Indicate the strategy followed in the conceptual modeling phase. Complete the schema documentation with any constraints not expressed by the conceptual ER schema. Do not report the full process but only the glossary, the skeleton schema and the final schema. **(7 points)**
- 2) Consider the conceptual schema defined in the previous exercise. Suppose that the following operations are carried out on this data:  
*Operation 1:* Insert a new article submission to a conference, including title, category, and the list of authors with their contact information – 50 times per day.  
*Operation 2:* For each program committee member, print the list of articles assigned for review, including article title, category, and the names of all authors – 40 times per day.  
*Operation 3:* Insert a new review for an article, including scores for originality, significance, quality, global score, and comments – 80 times per day.  
*Operation 4:* For each conference, print the list of all accepted articles grouped by category, including title, contact author name and email, and average global score – 10 times per day  
Taking into account that each conference has an average of 8 program committee members, and receives an average of 120 article submissions, each article has an average of 3 authors and 20% of articles are industrial papers, with each industrial paper having an average of 2 partners. Define the table of volumes and table of accesses for the conceptual schema, then design the logical schema of an object-relational database in UML and the corresponding types/table in SQL3 or Oracle. **(6 points)**
- 3) Given Operation 2 from Exercise 2, show a physical query plan for this query for a DBMS that supports primary heap data structures, secondary dense B+trees, and secondary hash indexes. Assume that currently no indexes have been created on any table. For each intermediate partial result, report the expected size and for each step, the expected cost. Hint: start from the result of algebraic optimization. **(6 points)**
- 4) Considering the database defined at Exercise 1, identify some facts, measures and dimensions for possible analysis, and define the schema of a Data Warehouse. **(4 points)**
- 5) Explain the warm restart procedure with an example. **(4 points)**
- 6) Explain the data replication techniques (and provide a taxonomy of them). Show pros/cons of every technique presented. **(3 points)**
- 7) Explain the substitution principle and how inheritance of typed tables is managed in SQL:2003 and Oracle. **(3 points)**