



UiT The Arctic University of Norway

| | |
|---|---------------------------|
| On-site exam in: | Inf-3701 |
| Date and time: | 14.05 2024, 09:00 – 13:00 |
| Course coordinator: | Randi Karlsen, 99297389 |
| Number of pages: | 2 including front page |
| Permitted aids: | None |
| Will course coordinator visit during the examination?: | No |
| Other: | |

Task 1: (30%)

- T1.1.** Assume a homogeneous distributed database. Describe how and why data partitioning is performed. Describe both horizontal and vertical partitioning.
- T1.2.** Describe what a *schemaless database* is. Describe also advantages and disadvantages of schemaless databases.
- T1.3.** Key-value store and Document store are two NoSQL data models. Describe both models and compare them. Include both similarities and differences.

Task 2: (20%)

- T2.1.** Describe the concept *interoperability*, including *syntactic* and *semantic interoperability*. Describe also the *challenges* related to semantic interoperability.
- T2.2.** Describe how semi-structured data can be used to overcome interoperability problems.

Task 3: (30%)

- T3.1.** What are web crawlers and how are they utilized?
- T3.2.** Describe what is OLAP, what is a data cube, and which are the 4 main operations within OLAP systems.
- T3.3.** Describe what is a data warehouse, illustrate its main architecture, and briefly describe the ETL tasks.

Task 4: (20%)

- T4.1.** Describe the challenges of false negatives and false positives occurring in document ranking strategies. Describe the two metrics to measure retrieval effectiveness, precision and recall.
- T4.2.** What is a classification problem? How are classification problems usually solved?