Data warehouses

Lecturer
Esteban ZIMANYI (Coordinator)

Course mnemonic
INFO-H419

ECTS credits
5 credits

Language(s) of instruction
English

Course period
First term

Course content

Relational and object-oriented databases are mainly suited for operational settings in which there are many small transactions querying and writing to the database. Consistency of the database (in the presence of potentially conflicting transactions) is of utmost importance. Much different is the situation in analytical processing where historical data is analyzed and aggregated in many different ways. Such queries differ significantly from the typical transactional queries in the relational model:

- Typically analytical queries touch a larger part of the database and last longer than the transactional queries;
- Analytical queries involve aggregations (min, max, avg, ...) over large subgroups of the data;
- When analyzing data it is convenient to see it as multi-dimensional.

For these reasons, data to be analyzed is typically collected into a data warehouse with Online Analytical Processing support. Online here refers to the fact that the answers to the queries should not take too long to be computed. Collecting the data is often referred to as Extract-Transform-Load (ELT). The data in the data warehouse needs to be organized in a way to enable the analytical queries to be executed efficiently. For the relational model star and snowflake schemes are popular designs. Next to OLAP on top of a relational database (ROLAP), also native OLAP solutions based on multidimensional structures (MOLAP) exist. In order to further improve query answering efficiency, some query results can already be materialized in the database, and new indexing techniques have been developed.

The first and largest part of the course covers the traditional data warehousing techniques. The main concepts of multidimensional databases are illustrated using the SQL Server tools. The second part of the course consists of advanced topics such as data warehousing appliances, data stream processing, data mining, and spatial-temporal data warehousing.

Objectives (and/or specific learning outcomes)

At the end of the course students are able to

- Understand the difference between operational databases and data warehouses
- Understand and be able to apply the principles of multidimensional modeling
- Exploit a data warehouse for querying and reporting
- Understand best practices and methodologies for data warehouse development
- Understand the process of populating a data warehouse from internal and external sources

Pre-requisits and co-requisits

Co-requisites courses

INFO-H415 | Advanced databases | 5 crédits, INFO-H417 | Database systems architecture | 5 crédits and INFO-H509 | Geo-Spatial and web technologies | 5 crédits

Teaching method and learning activities

- Theory lectures (24h) (including invited lectures by industrial partners)
- Exercises: pencil-and paper and lab sessions (24h)
- Group project (12h)

References, bibliography and recommended reading

Course books:

- Christian S. Jensen, Torben Bach Pedersen, Christian Thomsen. textit{Multidimensional Databases and Data Warehousing}. Morgan and Claypool Publishers. 2010
- Kimball, Ralph; Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker (2008). The Data Warehouse Lifecycle Toolkit (2nd ed.). Wiley.

Additional sources of information:

- Data Warehouse Design: Modern Principles and Methodologies, Golfarelli and Rizzi, McGraw-Hill, 2009
- Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications, Elzbieta Malinowski, Esteban Zimányi, Springer, 2008
- The Data Warehouse Toolkit, 2nd Ed., Kimball and Ross, Wiley, 2002

- Building the Data Warehouse. 4th edition. Inmon, Wiley, 2005
- Data Warehousing Fundamentals For IT Professionals. 2nd edition. Paulraj Ponniah, Wiley, 2010

Other information

Contact(s)

toon.calders@ulb.ac.be

Evaluation method(s)

Other

Evaluation method(s) (additional information)

- Written exam
- Group project

Determination of the mark (including the weighting of partial marks)

- Written exam (70%)
- Group project (30%)

Main language(s) of evaluation

English

Programmes

Programmes proposing this course at the Brussels School of Engineering

MA-IRIF | Master of science in Computer Science and Engineering | finalité Professional/unit 2 and finalité Big Data Management and Analytics (Erasmus Mundus)/unit 1

Programmes proposing this course at the faculty of Sciences

MA-INFO | Master in Computer science | finalité Professional/unit 2