Module Description, available in: EN

Data Management

General Information

Number of ECTS Credits

3

Module code

TSM_DataMgmt

Valid for academic year

2024-25

Last modification

2022-03-09

Coordinator of the module

Stefan Keller (OST, stefan.keller@ost.ch)

Explanations regarding the language definitions for each location:

- Instruction is given in the language defined below for each location/each time the module is held.
- Documentation is available in the languages defined below. Where documents are in several languages, the percentage distribution is shown (100% = all the documentation).
- The examination is available 100% in the languages shown for each location/each time it is held.

	Lausanne			Lugano	Zurich		
Instruction	**X** E 100%				**X** E 100%		
Documentation	**X** E 100%				**X** E 100%		
Examination	**X** E 100%				**X** E 100%		

Module Category

TSM Technical scientific module

Lessons

2 lecture periods and 1 tutorial period per week

Entry level competences

Prerequisites, previous knowledge

- UML Class Diagrams
- Relational Models, Relational Algebra
- Relational Database Management System (RDBMS) Architectures
- SQL:92 (i.e. queries with SELECT-FROM-WHERE-GROUP BY)
- Normalization, Query Optimization, Indexes
- Transaction Processing, Concurrency Control
- Security in Relational Database Systems

Brief course description of module objectives and content

This course is about Data Engineering and Information Retrieval. It covers methods and technologies for managing, processing and analyzing potentially large and distributed data collections for transactional or analytical use, including multi-model databases and NoSQL stores. And it covers also mastering data in unstructured form (full text search). The course consists of four parts: 1. Database Management; 2. Data Warehousing and Data Analytics (Business Intelligence); 3. Data Integration including Data Synthesis; and 4. Information Retrieval.

Aims, content, methods

Learning objectives and competencies to be acquired

This module covers following important aspects of Data Engineering:

- Students understand the use of modern database technologies for processing and managing potentially large and distributed data collections for transactional or analytical use.
- Students will be proficient in modern query languages such as the post-relational SQL (SQL:2023 and newer).
- Reaching beyond RDBMS, students learn about data structures (data types) and know which of these to use depending on the requirements and type of data available (polyglot persistence, multi-model databases).
- Students know NoSQL stores and selected cloud data stores.
- Students know methods and tools to integrate, to cleanse and to synthesize data.
- Students know how to deal with full text information using databases and search engines (information retrieval; prompt engineering).
- Students can also apply the acquired knowledge in their own working environment.

Module content with weighting of different components

The course is divided into four parts:

1. Database Management (DB): New data structures (types) and alternatives to RDBMS. Storing data with post- and non-relational aspects, including NoSQL technologies (especially graph databases), and a selection of advanced topics such as cloud or vector databases.
2. Data Warehousing and Data Analytics (DW): Methods and tools for data aggregation and data analytics such as the ones involved in business intelligence.
3. Data Integration (DI): Methods and tools for data integration, data cleansing and data synthesizing (e.g. for training and testing) are explained.
4. Information Retrieval (IR): Methods and tools for finding information in full text using databases and (enterprise) search engines, including crawling.

Weighting between the parts will be confirmed at the beginning of semester. Tentative weighting:

1. DB: ~4-6 weeks
2. DW: ~2-4 weeks
3. DI: ~1-3 weeks
4. IR: ~3-5 weeks

Teaching and learning methods

Frontal teaching, case studies, exercises, discussions, (group) work assignments (i.e. laboratory work or mini-project).

Literature

Optional literature suggestions (books):

- DB: Advanced Data Management for SQL, NoSQL, Cloud and Distributed Databases. R. Wiese. De Gruyter Textbook. 2015. ISBN 978-3-11-044140-6.
- DB: SQL for Data Scientists: A Beginner's Guide for Building Datasets for Analysis. R. Teate. Wiley. 2021. ISBN 978-1-119-66936-4.

- IR: Introduction to Information Retrieval. C.D. Manning, P. Raghavan, H. Schütze. Cambridge UP, 2008.
- IR: Information Retrieval in Practice. B. Croft, D. Metzler, T. Strohman. Pearson Education, 2009.

Assessment

Certification requirements

Module uses certification requirements

Certification requirements for final examinations (conditions for attestation)

A successful and timely submission of a lab report or a mini-project (with pass / no pass) may be required for admission to the exam. The exams may include topics from laboratory work or mini-projects.

Basic principle for exams

As a rule, all standard final exams are conducted in written form. For resit exams, lecturers will communicate the exam format (written/oral) together with the exam schedule.

Standard final exam for a module and written resit exam

Kind of exam

Written exam

Duration of exam

120 minutes

Permissible aids

Aids permitted as specified below:

Permissible electronic aids

Scientific calculator (without communication functions).

Other permissible aids

Summary on one A4 page (possibly written on both sides).

Special case: Resit exam as oral exam

Kind of exam

Oral exam

Duration of exam

30 minutes

Permissible aids

No aids permitted