Data Management

General Information

Number of ECTS Credits

3

Module code

TSM_DataMgmt

Valid for academic year

2021-22

Last modification

2021-02-12

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%		**X** E 100%		
Documentation	**X** E 100%				**X** E 100%		
Examination	**X** E 100%		**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

- Basic data structures and algorithms
- Working level on basic relational databases
- Relational Models, Relational Algebra,
- Normalization
- RDBMS architectures
- Transactions
- SQL:92
- Query optimization, Indexes
- Security in RDBMS

Brief course description of module objectives and content

This course is centered on the Data Engineering domain.

This course covers modern methods and technologies that are needed to manage and process potentially large, heterogeneous and distributed data collections. It includes diverse technologies frequently used in industrial contexts such as data warehouses, multi-model databases and NoSQL stores. A focus of the class is also given on Information Retrieval including techniques to efficiently retrieve data that are typically in unstructured form.

Aims, content, methods

Learning objectives and competencies to be acquired

Learning objectives and acquired competencies:

The learning objectives are directed towards Data Engineering:

- Students understand the modern ecosystems currently used in industries for data storage and data processing; including their respective adequations to application needs.
- Students understand the use of modern database and processing technologies for managing large, distributed and potentially heterogeneous data collections.
- Students are able to organize complex data structures, reaching beyond RDBMS and meeting the requirements of data availability and type, e.g. polyglot persistence and multi-model databases.
- Students are able to use selected advanced data technology stacks such as data warehouses, NoSQL stores and cloud data stores.
- Students are able to implement methods and tools to integrate, cleanse and synthesize data, such as the ones used to compose data pipelines.
- Students are able to integrate efficient Information Retrieval techniques typically used for unstructured and textual data, such as the ones used for search engines.
- Students can also apply the acquired knowledge in their own working environment.

Module content with weighting of different components

The module covers the following contents:

1. Database Management (DM): overview of modern ecosystems currently used in industries for data management; new data structures and alternatives to RDBMS; non-relational aspects including NoSQL and cloud data stores; new ways to query data such as JSON paths, SQL extensions, graph query language, etc.

2. Data Integration (DI): Data Warehousing for data aggregation and data preparation for analytics (e.g. business intelligence components); other methods and tools for data integration, data cleansing and data synthesizing.

3. Information Retrieval (IR): Efficient methods for finding information, typically in the context of unstructured and textual data, such as the ones used for search engines; ways to query data in IR systems, such as Query DSL.

Teaching and learning methods

Head-on teaching, exercises, case studies.

Literature

Optional literature suggestion (books):
- DB: Lena Wiese: Advanced Data Management for SQL, NoSQL, Cloud and Distributed Databases. De Gruyter Textbook. 2015. ISBN 978-3-11-044140-6.
- IR: Introduction to Information Retrieval. C.D. Manning, P. Raghavan, H. Schütze. Cambridge UP, 2008. Classical and web information retrieval systems: algorithms, mathematical foundations and practical issues.
- IR: Information Retrieval in Practice. B. Croft, D. Metzler, T. Strohman. Pearson Education, 2009.

Assessment

Module does not use certification requirements

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