Module	**Databases and Data Architecture Systems**
Code	MSLS_V5_4
Degree Program	Master of Science in Life Sciences (MSLS)
ECTS Credits	5
Workload	150h: 75h Lectures and Exercises, 75h Self-Study
Module Coordinator	**Name** Dr. Robert Vorburger **Phone** +41 (0)58 934 57 44 **Email** robert.vorburger@zhaw.ch **Address** ZHAW Zurich University of Applied Sciences Life Sciences and Facility Management Schloss 1 8820 Wädenswil
Lecturers	• Dr. Robert Vorburger • Adrian Busin
Entry Requirements	The course requires basic knowledge in the following topics: • Programming in Python • Statistical programming in R The scripting language Python as well as the statistical computing environment R are used in this module to create and process relational databases using SQL (structured query language). Prior knowledge of SQL is not required.
Learning Outcomes and Competences	Yes, it is true: *Data Scientist* is the sexiest job of the 21st century (at least according to the Harvard Business Review). While knowledge is usually engineered using statistical methods, the basis is always a well-structured set of data. The module covers the techniques and structures used to efficiently store, process, and load data in databases. By completing the module, students will specifically acquire knowledge and skills in the following fields: • Terminology and general basics of databases and data architecture systems • Different types of databases and their concepts • Relational databases and SQL • Python/R and SQL • Data Warehouses • NoSQL database concepts • Graph-based databases Hands-on exercises and examples will strengthen the student's competences in applying database concepts in the fields of life sciences.
Module Content	The module basically consists of four parts: • *Part I - Data and Data Architecture* ○ What is data?

	○ How is data stored and processed? ○ Databases vs file systems ○ Database-Management-Systems ○ Different types of databases: ▪ hierarchical ▪ network-oriented ▪ relational ▪ object-oriented • *Part II - Relational Databases* ○ Basic principles: entity integrity and referential integrity ○ Entity-Relationship-Model and Database Scheme ○ SQL (talk to the database) ○ Python and SQL ○ R and SQL • *Part III - Data Warehouse* ○ Extract-Transform-Load ○ OLAP-Cube ○ Business Intelligence • *Part IV* - NoSQL ○ Database types ▪ Key-Value-based ▪ Document-based ▪ Graph-based ○ Labeled Property Graph (including Neo4j and Cypher) ○ Triple Stores (including RDF and SPARQL) ○ Knowledge Graph
Teaching / Learning Methods	• Lectures : ~40% classical teaching / ~30% guided exercises • Self-Study : ~20% exercises / ~10% literature studying
Assessment of Learning Outcome	Programming assignments during the semester (20%) Final exam (written) (80%)
Bibliography	Important additional literature will be provided on Moodle.
Language	English
Comments	Data [ˈdeɪtə]: Borrowing from Latin *data*, nominative plural of *datum* (“that is given”), neuter past participle of dō (“I give”).
Last Update	12.09.2022