unibz

Fakultät für Informatik
Facoltà di Scienze e Tecnologie informatiche
Faculty of Computer Science

COURSE DESCRIPTION – ACADEMIC YEAR 2018/2019

Course title	**Programming Data Infrastructures**
Course code	73016
Scientific sector	INF/01
Degree	Master in Computational Data Science (LM-18)
Semester	2
Year	1
Credits	6
Modular	No

Total lecturing hours	40
Total lab hours	20
Attendance	Generally, attendance is not compulsory, but non-attending students have to contact the lecturer at the start of the course to agree on the modalities of the independent study.
Prerequisites	Elements of software engineering and programming, of the object-oriented paradigm (e.g., experience with JAVA), understanding of SQL
Course page	https://ole.unibz.it/

Specific educational objectives	The course belongs to the type "caratterizzanti – discipline informatiche" in the curriculum "Data Management". The course goal is designed to train students on how to organize and then process (big) data. To do so, students will learn how to write programming code that covers a mix of different techniques and data platforms for processing data that usually goes beyond traditional techniques (e.g., SQL). The course will be hands-on, that is, students will need to solve concrete problems using the techniques and tools introduced during the course. Then, students will learn how to discriminate which data processing techniques are suitable for particular problem depending of the context.

Lecturer	Ognjen Savkovic
Contact	BZ 2.02, ognjen.savkovic@unibz.it, +39 0471 016167
Scientific sector of lecturer	INF/01
Teaching language	English
Office hours	Arrange beforehand by email.
Lecturing Assistant (if any)	--
Contact LA	--
Office hours LA	--
List of topics	● Introduction to (big) data processing platforms ● Programming using Data Access Object patterns (e.g. Hibernate) ● Programming models for data analysis, (e.g., Hadoop) ● RDD-based programming (e.g., Spark) ● Parallel programming and functional style (e.g., Scala) ● Declarative languages for data processing (e.g., LogicBlox, DLV)
Teaching format	Frontal lectures and lab exercises. In the frontal lectures, the basic concepts are introduced and explained together with some examples. In the labs, the students will do a semester project, where selected

	techniques have to be applied to solve concrete problems and provide software solution.

Learning outcomes	Knowledge and understanding: • D1.2 - Understanding of the skills, tools and techniques required for an effective use of data science • D1.3 - Knowledge of principles, methods and techniques for processing data in order to make them usable for practical purposes, and understanding of the challenges in this field • D1.4 - Sound basic knowledge of storing, querying and managing large amounts of data and the associated languages, tools and systems • D1.5 - Knowledge of principles and models for the representation, management and processing of complex and heterogeneous data Applying knowledge and understanding: • D2.1 - Practical application and evaluation of tools and techniques in the field of data science • D2.4 - Ability to develop programmes and use tools for the analysis and management of data and related infrastructures • D2.10 - Application of languages, tools, and methods for the design of information systems and their corresponding software applications for data, process, and organization management Making judgments • D3.1 - Ability to plan and, if necessary, re-plan a technical project activity for the analysis and management of data, or for the implementation of corresponding software systems or applications, and to complete it within the defined deadlines • D3.2 - Ability to autonomously select the documentation (in the form of books, web, magazines, etc.) needed to keep up to date in a given sector • D3.3 - Ability to identify reasonable work goals and estimate the resources needed to achieve these goals. Communication skills • D4.1 - Ability to use English at an advanced level with particular reference to disciplinary terminology • D4.3 - Ability to structure and draft scientific and technical documentation Learning skills • D5.1 – Ability to autonomously extend the knowledge acquired during the study course • D5.3 – Ability to deal with problems in a systematic and creative way and to appropriate problem solving techniques

Assessment	The assessment of the course consists of two parts: • several assignments which are done during the semester and require students to solve exercises by using methods and techniques; • a final project which is done during the semester and requires students to solve a concrete problem by applying methods and

	• techniques taught in the course (80% of the mark). The final project is delivered in the form of a project report, the developed code, and a short oral presentation. The exam modalities are the same for attending and non-attending students.
Assessment language	English
Assessment Typology	Monocratic
Evaluation criteria and criteria for awarding marks	The final exam grade is the weighted average of the assignment marks (20%) and the final project (80%). Criteria for the evaluation of the assignments: correctness of the solution. Criteria for the evaluation of the project: technical complexity of the solution, originality and suitability of the idea.

Required readings	The exact sources will be announced during the course. Subject Librarian: David Gebhardi, David.Gebhardi@unibz.it
Supplementary readings	Students can prepare better before and during the course by reading some of the following material: • Online: http://hibernate.org/orm/ • Book: Professional Hadoop. Benoy Antony et al. Series: Wrox Programmer to Programmer (free online from unibz Library) • Book: Spark. Ganelin, Ilya. (free online from unibz Library) • Book: Scala cookbook. Alvin Alexander 2013. (free online from unibz Library) • Book: Logic programming with Prolog. Max Bramer (free online from unibz Library)
Software used	Java, Hibernate, Hadoop (MapReduce) framework, Spark, Scala, Logic Programming Tools (Prolog, DLV)