COURSE DESCRIPTION – ACADEMIC YEAR 2018/2019

Course title	**Real-Time Big Data Processing**
Course code	73033
Scientific sector	INF/01
Degree	Master in Computational Data Science (LM-18)
Semester	2
Year	1 and 2
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	
Course page	https://ole.unibz.it/

Specific educational objectives	The course belongs to the type "caratterizzanti – discipline informatiche" in the curricula "Data Analytics" and "Data Management". The course aims at teaching both scientific foundations and practical aspects of real-time big data processing technologies. The students will learn the basic concepts of such systems and how to use them to solve concrete problems. Moreover, students will be trained to evaluate the advantages and disadvantages of such technologies in different application contexts.

Lecturer	Guohui Xiao, http://www.ghxiao.org
Contact	Piazza Domenicani, 3, Room 205, xiao@inf.unibz.it, 0471 016 267
Scientific sector of lecturer	ING-INF/05
Teaching language	English
Office hours	arrange beforehand by email.
Lecturing Assistant (if any)	--
Contact LA	--
Office hours LA	--
List of topics	• Complex event processing • Stream data mining • Semantic techniques for Streaming data • Programming models for streaming data • CEP programming • Distributed real-time computation system (e.g., Apache, Flink, Storm) • Functional reactive programming • Robustness of stream processing
Teaching format	Frontal lectures and project work during the exercise hours. 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.

Learning outcomes	Applying knowledge and understanding: • D1.1 - Knowledge of the key concepts and technologies of data science disciplines • 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 • D2.1 - Practical application and evaluation of tools and techniques in the field of data science • D2.2 - Ability to address and solve a problem using scientific methods Making judgments • 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 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 course of study • D5.2 - Ability to autonomously keep oneself up to date with the developments of the most important areas of data science • 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 is based on a project which is done during the semester and requires students to solve a concrete problem by using methods and technologies taught in the course (100% of the mark). The project verifies whether the student is able to apply advanced data management techniques to solve concrete problems. The project is assessed through a final presentation, demo and project report. 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 project mark (100%). Criteria for the evaluation of the project: correctness of the solution, complexity of the project, technologies used in the solution, quality of the report and the presentation.

Required readings	There is no single textbook that covers the entire course. The course material is collected from various textbooks and research papers, including the following ones: • Martin Kleppmann: Designing Data-Intensive Applications - The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. O'Reilly, 2016. • Ellen Friedman & Kostas Tzoumas: Introduction to Apache Flink - Stream Processing for Real Time and Beyond. O'Reilly, 2017. • Stephen Blackheath and Anthony Jones: Functional Reactive Programming. Manning, 2016. Subject Librarian: David Gebhardi, David.Gebhardi@unibz.it
Supplementary readings	Additional sources will be announced during the course.
Software used	Languages: Python, Java, Scala Software: Apache Kafka, Apache Spark, Apache Flink, Apach Beam, Apache Casandra, ReactiveX