Title: **From Big Data to Big Multidimensional Data: Models, Issues, Challenges**
Web site: https://sites.google.com/unical.it/big-multid-data-prof-cuzzocrea/home

Speaker: **Prof. Alfredo Cuzzocrea**
iDEA Lab, Head
University of Calabria, Italy
Excellence Chair in Computer Engineering
LORIA, France
web: http://si.dimes.unical.it/cuzzocrea

Abstract and Lecture Summary: Big data is gaining momentum in the research community, due to the several challenges posed by the management of such kind of data. Big data are relevant not only in the academic context, but also in the industrial context, where they play the major role. Indeed, several kinds of application are now exploiting big data, such as: Web advertisement, social network intelligence, e-science applications, smart city applications, and so forth. Among big data, big multidimensional data are a special case of big data that fully expose the "famous" 3V (volume, velocity, variety) and are of relevant interest at now. In this course, tailored to PhD students in Computer Science and Computer Engineering, we first investigate foundations of big data, critical state-of-the-art analysis, research challenges and industrial applications. After that, we move the attention on special lectures focused on big multidimensional data.

Lecture 1: Big Multidimensional Data
Summary: Lecture 1 details on big multidimensional data, a special kind of big data where the main emphasis is on the multidimensionality of data, still influenced by the long tradition of OLAP and Data Warehousing methodologies developed during past decades. The lecture focuses on foundations and applications of big multidimensional data, by analyzing similarities and differences with past research efforts. State-of-the-art proposals are also reported and critically discussed. Finally, several alternatives for representing and managing big multidimensional data are presented in detail.

Lecture 2: Supporting Compression and Accuracy of Big Multidimensional Data
Summary: Lecture 2 introduces some techniques for supporting compression and accuracy of big multidimensional data, with particular regards to OLAP data cubes. The proposed techniques can be efficiently used in Quality-of-Answer-based OLAP tools, where OLAP users/applications and Data Warehouse servers are allowed to mediate on the compression and accuracy of (approximate) answers. Two techniques are presented: D-Syn, which exploits an analytical interpretation of data cubes for making the intrinsic data cube compression more flexible, and LCS-Hist, which is able to deal with high-dimensional data cubes.

Lecture 3: Aggregating Big Multidimensional Data Streams
Summary: Lecture 3 highlights the streaming nature of big multidimensional data, and puts the attention on the problem of effectively and efficiently aggregating big multidimensional data streams. Two methodologies are presented. The first one, called CAMS, is capable to deal with the special features of such kind of data, and computing aggregations based on a complex approach that combines several intelligent techniques for processing streaming data. The second one, called nlMRDS, allows us to achieve higher performance, thanks to fortunate compression paradigms.

Lecture 4: OLAPing Uncertain Big Multidimensional Data Streams
Summary: Big multidimensional data streams are playing a leading role in next-generation DSMS. This essentially because real-life big data streams are inherently multidimensional, multi-level and multi-granular in nature, hence opening the door to a wide spectrum of applications ranging from environmental sensor networks to monitoring and tracking systems, and so forth. Consequently, there is a need for innovative models and algorithms for representing and processing such streams. Moreover, supporting OLAP analysis and mining tasks is a "first-class" issue in the major context of knowledge discovery from streams, for which above-mentioned models and algorithms are baseline components. This issue becomes more problematic when uncertain and imprecise multidimensional big data streams are considered. Inspired by these critical research challenges, Lecture 4 presents an overview of major research issues in this context and an innovative technique for supporting OLAP over uncertain big multidimensional data streams.

Lecture 5: Privacy-Preserving Big Multidimensional Data Management
Summary: The problem of making privacy-preserving big multidimensional data is a major research topic in Big Data research. Here, several approaches have been proposed and explored in different computational settings. Lecture 5 will discuss the fundamental of privacy-preserving OLAP and proposes a novel technique used to obtain privacy-preserving data cubes, a special case of big multidimensional data, that "balance" accuracy and privacy constraints for a wide family of next-generation applications, also falling in the modern context of Cloud Computing. The proposed technique is based on nice flexible sampling-based techniques, which allow us to gain in "query coverage" while spending something in terms of computational overheads. Experimental result analysis will be provided, with discussions of trade-offs and pro/cons assessment.

Course Duration
15 hours

Course Prerequisites
Big data foundations

Examination and Grading
Oral presentation of a course's topic, plus general examination on all the course's topics. At the end of the exam's successfully-pass, an official certificate will be released. This certificate will be used with the UniBo PhD Office for proofing the achieved credits.

Course Calendar
February 17, 2021 – 9:30-12:30 (Microsoft TEAMS)
February 18, 2021 – 9:30-12:30 (Microsoft TEAMS)
February 19, 2021 – 9:30-12:30 (Microsoft TEAMS)
February 22, 2021 – 9:30-12:30 (Microsoft TEAMS)
February 23, 2021 – 9:30-12:30 (Microsoft TEAMS)

Examination Days (Tentative)
March 10, 2021 – 9:30-12:30 (Microsoft TEAMS)
March 11, 2021 – 9:30-12:30 (Microsoft TEAMS)

Course Material
Delivered on Microsoft TEAMS