

A Multidisciplinary Perspective On Big Data 2017

Last Updated Tuesday, 20 June 2017

This is the official web page of the 2017 edition of the PhD course of Politecnico di Milano on "A Multidisciplinary Perspective On Big Data".

LECTURERS

Responsible : Emanuele Della Valle

Lectures: Danilo Ardagna, Michela Arnaboldi, Paolo Ciuccarelli, Emanuele Della Valle, Simone Vantini

Other Lecturers: Cinzia Cappiello, Paolo Cremonesi, Elisabetta Di Nitto, Pier Luca Lanzi, Letizia Tanca

MISSION AND GOALS

The term Big Data refers to a growing torrent of information that, if successfully analyzed, can unleash new business opportunities and revenues. This course aims at introducing Big Data analytics methods and includes practical sessions on PoliMI's Big Data computational infrastructure.

CLASSES

- Introduction to big data and data base fundamentals (3 hours)
 - 1.3.2017 - 10:00-13:00 - in the Seminar Room of building 20
prof. Della Valle and prof. Tanca
 - What is Big Data? volume, velocity, variety, veracity, … to create value [slides from 1 to 19]
 - data base fundamentals [slides]
 - Paradigm shifts enabled [slides from 20 on]
- Mastering the volume dimension (6 hours)
 - 8.3.2017 - 9:30-11:30 in the Conference Room of building 20
prof. Cremonesi
 - Introduction to cloud computing and Technologies for Infrastructure-as-a-Service [slides]
 - 8.3.2017 - 14:30-16:30 in the Conference Room of building 20
prof. Ardagna
 - Map Reduce from Hadoop to Spark [slides]
- 9.3.2017 - 10:00-12:00 in the Seminar Room of building 20
prof. Di Nitto
 - NoSQL databases for Big Data [slides]
- Mastering the variety dimension (3 hours)
 - 14.3.2017 - 10:00-11:00 in the Conference Room of building 20
prof. Tanca
 - Data Integration principles, approaches and tools [slides]
 - 14.3.2017 - 11:00-13:00 in the Conference Room of building 20
prof. Della Valle and
 - The role of Ontologies and Semantic Web technologies [slides]
- Mastering the velocity dimension (2 hours)
 - 16.3.2017 - 10:00-12:00 in the Conference Room of building 20
prof. Della Valle
 - Information flow processing principle, approaches and tools [slides]
- velocity for Big Data with Spark [slides][video]
- Mastering the veracity dimension (2 hours)
 - 16.3.2017 - 14:30-16:30 in the Conference Room of building 20
prof. Cappiello [slides]

- data quality, definitions, dimensions, approaches and tools
- uncertainty and data quality problems in big data
- Making sense of Big Data (8 hours)
- 21.3.2017 - 10:00-12:00 in the Seminar Room of building 20
prof. Ciuccarelli
- Visual analytics in the Big Data era [slides !!!113MB!!!]

- 22.3.2017 - 10:00-12:00 in the Conference Room of building 20
prof. Lanzi
- Knowledge discovery and Data Mining in the Big Data era [slides]

- 23.3.2017 - 10:00-13:00 in the Conference Room of building 20
prof. Vantini
- Big Data Analytics [slides]
- Creating Value with Big Data (3 hours)
- 21.7.2017 - 10:00-13:00 in the Seminar Room of building 20
prof. Deborah Agostino [slides, excel]
- Student reporting session (4 hours or less)
- 11.7.2017 - 14:00-17:00 - in the PT1 room of building 20

STUDENT REPORTING

Students are expected to form multidisciplinary teams and work together on a project. Please, send an email to emanuele.dellavalle - at - polimi.it in order to get your project approved.

The email should says:

- who are the members of the team (put emphasis on the multidisciplinary composition)
- what problem you want to solve
- what data you plan to work on
- why this qualifies as a Big Data problem

The final presentation, in addition to the basic information above, should also include

- how you addressed the problem from a methodological and practical perspective
- what you successfully completed (a demo is very appreciated)
- what you failed to complete and why
- discussion on lesson learnt (emphasis on multidisciplinary is appreciated)

Ideally, the presentation for a group of 3 people should last 20-25 minutes. All team members should speak. All team members should know in deep a part of the work, but also to some level of details the work done by the others.

TEACHING MATERIALS

The course material consists in slides prepared by the lecturers, links to on-line tutorials, and the dataset of Telecom Italia Big Data Challenge.