

# A Multidisciplinary Perspective On Big Data 2015

Last Updated Tuesday, 28 February 2017

This is the official web page of the 2015 edition of the PhD course of Politecnico di Milano on "A Multidisciplinary Perspective On Big Data". If you are a student of the 2017 edition, this is the official page. Lecturers

Responsible : Emanuele Della Valle Lectures: Danilo Ardagna, Cinzia Cappiello, Paolo Ciuccarelli, Paolo Cremonesi, Emanuele Della Valle, Elisabetta Di Nitto, and Letizia Tanca

Invited Speakers: Fabrizio Antonelli (Telecom Italia), Stefano Ceri, Piero Fraternali, Marco Monti (IBM), and Marco Tagliasacchi

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

- 1) Introduction to big data - Emanuele Della Valle (2 hours) [pdf]
  - Why now?
  - What is Big Data? volume, velocity, variety, veracity, &hellip;
  - Paradigm shifts enabled
  - Tools and Market Landscape
- 2) Infrastructures for big data problems to master the volume dimension - Paolo Cremonesi & Danilo Ardagna (4 hours)
  - Introduction to cloud computing and Technologies for Infrastructure-as-a-Service [pdf1,pdf2]
  - Map Reduce, Hadop, Hadoop ecosystem and Map Reduce Cloud based solutions [pdf]
  - BigSheet and BigSQL tools [pdf]
- 3) Integrating and analyzing massive and heterogeneous (relational, graph, semi-structured) data collections and streams - Elisabetta Di Nitto, Letizia Tanca, and Emanuele Della Valle (4 hours)
  - Overview of NoSQL databases and Big data and NoSQL [pdf]
  - mastering the variety dimension [pdf]
  - mastering the velocity dimension [pdf]
- 4) mastering the veracity dimension - Cinzia Cappiello (3 hours) [pdf]
  - data quality
  - definition and dimensions
  - techniques for data quality assessment and improvement
  - uncertainty and data quality problems in big data"
- 5) Making sense of Big Data - Paolo Ciuccarelli, Letizia Tanca, and Paolo Cremonesi (6 hours)
  - Big Data Visualisation [pdf]
  - Big Data Exploration, summarisation and context-aware reduction [pdf]
  - Recommender Systems [pdf]
- 6) Open Workshop on Big Data (4 hours) (storify)
  - genomics applications - Stefano Ceri [pdf]
  - cognitive computing - Marco Monti (IBM)
  - multimedia analytics - Marco Tagliasacchi [pdf]
  - Telecom Italia Big Data Challenge - Fabrizio Antonelli (Telecom Italia) [pdf]
  - round table
- 7) Putting it all together on Telecom Italia Big Data Challenge dataset - Students (4 hours)
  - Assignments to students
  - Students's reports Calendar

Please find hereafter the current version of the calendar. Missing times will appear shortly. The open workshop (part 6) will most likely take place the last week of January. The students presentations (part 7) will take place in the last week of February. DayTimeAula LecturerPart

Title

1.12.2014 10:00-12:00 PT1Emanuele Della Valle

1

Introduction to big data [pdf] course introduction [pdf]

5.12.2014 9:30-12:30 Sala Conferenze Paolo Cremonesi

2

Infrastructures for big data problems to master the volume dimension [pdf1, pdf2]

5.12.2014 14:30-16:30 Sala Conferenze Danilo Ardagna 2 Infrastructures for big data problems to master the volume dimension [pdf]

13.1.2015 9:00-12:00

Sala Conferenze Elisabetta Di Nitto 3 Overview of NoSQL databases and Big data [pdf] 13.1.2015 13:30-14:30 Sala Conferenze

Filippo Mariani 2 BigSheet and BigSQL tools [pdf]

14.1.2015 14:30-15:30

Sala Conferenze Letizia Tanca

3 Mastering the variety dimension [pdf]

14.1.2015 15:45-16:45 Sala Conferenze Emanuele Della Valle

3 Mastering the velocity dimension [pdf]

16.1.2015 9:30-12:30 Sala Seminari Cinzia Cappiello 4 Mastering the veracity dimension [pdf] 20.1.2015 10:00-

11:00 Sala Conferenze Paolo Ciuccarelli

5 Big Data Visualisation [pdf]

20.1.2015 11:00-12:00

14:00-16:00

Sala Conferenze Letizia Tanca

5 Big Data Exploration, summarisation and context-aware reduction [pdf] 20.1.2015 16:00-17:00 Sala Conferenze Paolo

Cremonesi 5 Recommender Systems [pdf]

27.1.2015 14:30-17:30 Sala Conferenze Emanuele Della Valle

6

Open workshop on Application of Big Data (flyer [pdf])

- Fabrizio Antonelli: Telecom Italia Big Data Challenge 2014: dataset, applicants, and winners [pdf]

- Stefano Ceri: Genomic Computing in the Big Data era [pdf]

- Marco Monti: Managing Variety of Big Data in the Healthcare and Financial domain [pdf]

- Marco Tagliasacchi: Analysis of Big Visual Data [pdf]

27.2.2015 9:30-13:30 Sala Seminari

Emanuele Della Valle

4

Students' reports

- Angela Geronazzo, Giovanni Meroni, Riccardo Pinciroli, and Giovanni Quattrocchi: Characterizing news according to heterogeneous communication data variations

- Chiara Tagliaro, Claudio Menghi, Alessandro Rizzi e Marco Scavuzzo: VOL&rsquo;s UP

- Mehrnoosh Askarpoura and Narges Shahmandi hoonejani: Performance evaluation of a Spark Cluster using Facebook Social Network as a Case Study

- Marco Miglierina, Giovanni Paolo Gibilisco, Marzia Bolpagni, Maristella Lucchini: Do they care? The social side of the news

- Mona Nasery, Roberto Pagano, Massimo Quadrana: Visualize call traffic intensity between any selected zones in Milan Teaching material

The course material consists in slides prepared by the lecturers, links to on-line tutorials, and the open source release of the dataset of Telecom Italia Big Data Challenge. Students will gain enough background on the topics to be able to use the infrastructure made available by IBM as well as the poliCloud one donated by Yahoo!. Giving the exam The exam consist in:

- formulating a Big Data problem: is it about volume? velocity? variety? veracity? a mix of them? why?

- solving such a problem with Big Data technologies illustrated during the course: which tools? Which methods? Why?

- preparing a presentation where you discuss:

- who you are (name, surname, skills, time spent)

- the problem

- the (partial) solution

- what you learnt, i.e., what did go as expected? what did not? why? (right/wrong skills/tools, lack of time, etc.)

- giving your presentation
- discuss the presentations of others

Students are expected to cluster in groups of maximum 4 people, but one can work on his/her own if he/she prefers. Please find at the following link a form to fill up:

<https://docs.google.com/forms/d/18vizeCbt98017IXrIA7aQWTeSFdBxQR5ONnUGvLQyY/viewform>