



POLITECNICO
MILANO 1863

SYSTEMS AND METHODS FOR BIG AND UNSTRUCTURED DATA

Marco Brambilla

marco.brambilla@polimi.it

 @marcobrambi

The team

- Main Instructor: Marco Brambilla

 <http://home.dei.polimi.it/mbrambil>

 marco.brambilla@polimi.it

 [marcobrambi](https://www.linkedin.com/in/marcobrambi)

 [marcobrambi](https://twitter.com/marcobrambi)

 slideshare.com/mbrambil

- Assistant: Andrea Tocchetti

 Andrea.tocchetti@polimi.it

Course Focus

A **broad vision** of

- Modern enterprise data management **systems**
- **Models of data**
- **Languages of data management**
- **Technologies and platforms/architectures** that implement them
- **Methods** that support design of solutions

Course Agenda – 1 – BIG DATA

1. Approaches to Big Data management

- Big Data problems and dimensions
- Data engineering and data science pipeline
- Enterprise-scale data management
- Scalability and persistency vs. volatility issues
- Cross-source data integration problems and architectures
- CAP theorem and implications. Non-relational distribution architectures
- Evolution of transactional properties: from ACID to BASE. Modern transactional architectures
- Data sharding and replication

Course Agenda - 2 - SYSTEMS

2. Systems and Models for Big and Unstructured Data

- Graph databases
- Semantic databases
- Columnar databases
- Document-oriented databases
- Key-value databases
- IR-based databases
- Vector databases
- Feature stores

Course Agenda - 2 - SYSTEMS

- Each category of systems is covered along 5 dimensions:

(1) data model

(2) (declarative vs. imperative) query languages

(3) data distribution

(4) non-functional aspects

(5) architectural solutions

Course Agenda - 3 – METHODS

3. Methods for the Design of Applications

- Modeling languages and methods for building unstructured data applications
- Design methodology within the data engineering pipeline
- Schema-less, implicit-schema, and schema-on-read approaches

Course Modalities

- Frontal teaching
- Technology hands-on
- Online materials & readings
- Workgroup project

Course Evaluation

Constraint: sufficient mark = 60% for each part and 18 points total

Project Work

- Groups of max. 3 people
- Selection of one project topic
- Further details later

Ongoing Evaluation

- 3 very simple assignments along the course (to be solved in class)

Materials

- Slides, readings, book(s)
- On BEEP platform



Bibliography

Marting Kleppmann, **Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems**,
Editore: O'Reilly, ISBN: 978-1449373320

Slides

Course notes – book in preparation

References

- <http://home.deib.polimi.it/mbrambil>
- <http://datascience.deib.polimi.it/>

WeBeep

Online calendar



POLITECNICO
MILANO 1863

THANKS! QUESTIONS?

Marco Brambilla @marcobrambi marco.brambilla@polimi.it
<http://datascience.deib.polimi.it> <http://home.deib.polimi.it/marcobrambi>