POLITECNICO
MILANO 1863

SYSTEMS AND METHODS FOR BIG AND UNSTRUCTURED DATA

Project Work 2022

Marco Brambilla

marco.brambilla@polimi.it

@marcobrambi

Project Work

Group work
Incremental deliveries

Course Evaluation

30% group project delivery

20% closed-book, multi-choice questionnaire

50% written exam, open questions and exercises

Project Work

Groups of max. 5 people

Selection of one project topic

Data and architecture design and implementation

Activity

Specification of problem
Data design
Graph
Document
Spark

Topic Description

Bibliography Database

Scientific articles

Authors, affiliations

Journals and conferences

DOI

References

Dblp.org

full XML dump: https://dblp.uni-trier.de/xml/

DOI.org

ORCID.org

+ abstracts and further metadata
  - keywords
  - ...

Deliveries

Delivery #1

ER model
Graph DB
Data upload (importing)
Data queries

Cypher Queries for Delivery #1

Minimum 5 (diverse) data creation/update commands
Minimum 10 queries
Minimum complexity of queries:

- 3 nodes, conditions
- 3 nodes, conditions, aggregation
- 5 nodes, conditions, aggregations, limits
- Functions (minimum path)

Check complexity / performance time

THANKS!
QUESTIONS?

POLITECNICO
MILANO 1863

Marco Brambilla @marcobrambi marco.brambilla@polimi.it

http://datascience.deib.polimi.it http://home.deib.polimi.it/marcobrambi