



Project Work

Marco Brambilla

@marcobrambi

marco.brambilla@usi.ch

Project Work

- Group work
- No team changes allowed
- 1 assignment
- Maximum 30% (3 pts. Total) on the final mark
 - 1 point for each project

Activity

Specification of problem of choice

Dataset identification (search or implement)

Data design

Implementation based on one technology per project, compatible with the scenario and data needs:

- Neo4J
- MongoDB
- Spark

Presentation-shaped report

Public presentation in class

Groups Registration

Link to Google spreadsheet:

<https://docs.google.com/spreadsheets/d/1JMTAhf6247GtnMgKc9oW0Hd8HPkLzFY4vA7nceTERio/edit?usp=sharing>

- Add all members
- Add a topic for the project work

Task

- Define the project scenario
- Identify a dataset
 - Pre-existing (open data, from Dataverse, Zenodo, or similar)
 - Random
 - Created by your team
- Run the importing process
- Draw a ER model of the data
- Draw a “schema” or example instances in the NoSQL model
- List the size of the different dataset items (how many instances)
- Run queries
- Add your notes / critical assessment
- Prepare a presentation

Example: AirBNB

- Import the AirBnB dataset
- Download the listings and reviews datasets
- Import the data
- Implement the query
- Opt.: Build the AirBNB web site
- Opt.: Become a billionaire 😊

Delivery list

- Scenario description
- ER model
- NoSQL model
- Implementation
 - Data structure and instances
 - Minimum 12 queries, 5 editing commands
- Short report describing the work done and your comments/conclusions
- Optional: application (web?) implementation besides the queries in the tool
- Presentation to be shown in class + demo (presentations: rotation)

Queries for Delivery

Diverse data creation/update commands representing typical real-world changes and queries on the data

Diverse queries showing real-world data needs, e.g. grouping, searching, counting, combining, aggregating, c

In general, make sure the queries are reasonably complex

Check performance time and comment on the results

Note: query complexity and usefulness will be rewarded



Università
della
Svizzera
italiana



Data Design and Modeling course

Marco Brambilla

@marcobrambi

marco.brambilla@usi.ch