Project
Implementation of a (Big) Data Management Backbone

Organization

Big Data Management – FIB – UPC

Organization

Teams - Local Masters students

- Work in pairs
  - You have to define the teams
  - All pairs must be different in P1 and P2
- How to deal with the incremental nature of the project?
  - You are free to extend the solution of the other team member
    - As long as both members of the team agree
  - Otherwise, use the provided solution

Teams - Erasmus Mundus students

- Work with your business group
- Large teams (>3 people) will be split
  - 2+2
- The configuration of each sub-team will be made by the lecturers

Development environment

- Virtual machines hosted at FIB
  - Ubuntu Desktop with HDFS, HBase and MongoDB installed in standalone mode
  - See the manual in LearnSQL
  - Credentials will be provided in the team's description
- Your own development environment
  - Java (intellij IDEA)
  - Python (PyCharm)

Validation tests - Local Masters students

- Each part (P1, P2) will have associated a validation test
  - See specific dates in LearnSQL
- Individual test
  - Questions related to the project development and its relationship with the concepts studied in class

Evaluation - Local Masters students

Final Mark = min(10 ; 60%E + 40%L + 10%P)

L = Weighted average of the marks of the three lab deliverables

E = Final exam

P = Participation in the class

- L = (1/3) * P1 + (2/3) * P2
- Where, each Pi is computed as
  - Pi = 0,4 * Ti + 0,6 * Di
  - where Ti is the mark on the validation test, and Di is the deliverable's mark

Evaluation - Erasmus Mundus students

Final Mark = min(10 ; 60%E + 40%L + 10%P)

L = Weighted average of the marks of the three lab deliverables
E = Final exam
P = Participation in the class

- L = (1/4) * P1 + (1/2) * P2 + (1/4) * Pres.
- Where, each Pi is the deliverable's mark and Pres. is the final presentation's mark

Deliverables

- Document (max 5 pages)
  - Describe all relevant choices and specificities of your project
  - Justify choice of modeling approach and technology
    - These must be supported by the concepts studied in class
  - Present in a high-level manner (BPMN, sequence diagrams, boxes and arrows, …) the data transformations implemented
  - Remember, there is not a single correct solution
    - The most important part is how you **justify your choices**, and **discuss pros/cons**
- All code required to deploy the proposed design
  - Java/Python
  - Scripts

Closing