Big Data

Modern Data Management Systems

Themis Palpanas

Paris Descartes University
Institut Universitaire de France

1

LIPADE Laboratoire d'Informatique PAris DEscartes

diNo

1

Organization

- Instructor
  - Themis Palpanas: themis@mi.parisdescartes.fr
- Teaching Assistant
  - N/A
- Office Hours
  - upon request; after lecture (preferable)
- Structure + Grades
  - Classes
  - Assignments (30%): paper presentation
  - Group Projects (50%): in groups
  - Exam (20%): oral examination on project
- Course Web Page:
  - N/A

2

2

Project

- Carry out in groups of 4 students
- Make a proposal after two weeks
- Submit one week before the exam
  - written report
  - code
- two alternatives:
  - 1. experimental
  - 2. comparative evaluation

3

3

Project

- Experimental
  - define a problem you want to solve
  - read relevant published papers
    - search papers using „complete search“ of DBLP:
      - http://dblp.uni-trier.de/search/publ
  - propose and describe a solution
    - may be based on previous work/ideas
    - proposing something new/better gives you bonus marks!
  - implement your solution
    - write code
    - run experiments
  - write 10-page report
    - discuss the problem, your solution, your implementation and your results
- ideas
  - analyze social media data
  - ...

4

4

Project

- Comparative Evaluation
  - read published papers on some specific problem
    - search papers using „complete search“ of DBLP:
      - http://dblp.uni-trier.de/search/publ
  - implement 3 of the solutions (for the same problem)
    - write code
    - run experiments
    - compare the results
  - write 10-page report
    - discuss the problem, the solutions, your implementation and your results
- ideas
  - graph databases
  - join algorithms in map-reduce
  - …

5

5

Logistics

- Course
  - Thu 12:45-2pm
- TD
  - Wed 1-2:15pm
  - Wed 2:30-3:45pm

7

7

Syllabus Overview

- advanced topics on relational databases
- non-relational data management systems
- data management on new hardware

8

Pre-requisites

- Good knowledge of relational databases
  - ER modeling
  - Relational databases concepts
- SQL programming
- Another high level programming language
- Familiarity with basic system administration

9

What you will achieve

- Understand the theory and process of data management
- Gain knowledge of the new data management trends
- Be able to design the solution for a data management problem

10

References

- course slides
- books
  - Mining of Massive Datasets, by Jure Leskovec, Anand Rajaraman, and Jeff Ullman
    - http://infolab.stanford.edu/~ullman/mmds/book.pdf
  - Database Management Systems, by Raghu Ramakrishnan and Johannes Gehrke
    - https://cs2102-intro-to-database.googlecode.com/files/Database%20Management%20Systems%203Rd%20Edition.pdf
  - Database Systems: The Complete Book (DS:CB), by Hector Garcia-Molina, Jeff Ullman, and Jennifer Widom
    - http://infolab.stanford.edu/~ullman/dscb.html
  - The Fourth Paradigm: Data Intensive Scientific Discovery, Edited by Tony Hey, Stewart Tansley, and Kristin Tolle
    - http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf
  - ...

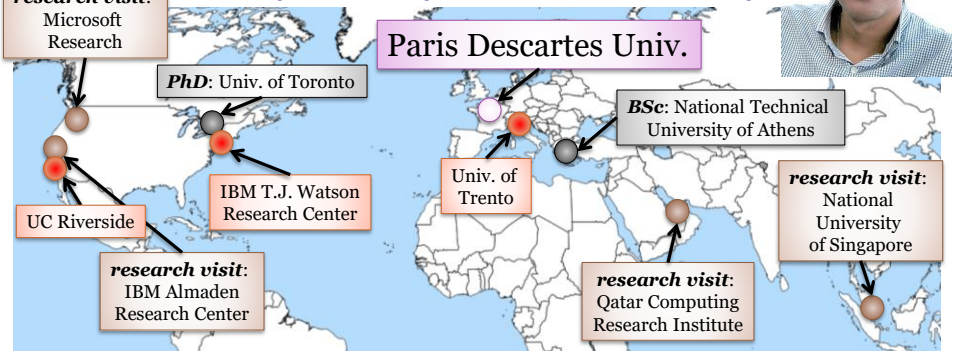
11

who is who

12

Themis Palpanas

http://www.mi.parisdescartes.fr/~themisp



- **Senior Member** of the **Institut Universitaire de France (IUF)**
- **director of database group (diNo)** of LIPADE, Paris Descartes University
- founder and **director of database group (dbTrento)** of University of Trento
- **awards**: IBM SUR Award, 3 Best Paper Awards
- **9 US patents**: 3 implemented in IBM **commercial products**
- **expertise in databases, data mining, data sequences**

Themis Palpanas - January 2018

13

Our team!

- if you are interested in this area
  - find out more
  - and contact us:

- http://www.mi.parisdescartes.fr/~themisp/
- http://dino.mi.parisdescartes.fr/