

BDA THEMATIC SCHOOL – DISTRIBUTED MASSIVE DATA

“Scalable Data Analytics”

Aussois, 17 – 22 June 2018

Luc Bouganim

Directeur de Recherche, INRIA, luc.boufanim@inria.fr

Cédric du Mouza

Maître de Conférences, CNAM, dumouza@cnam.fr

Genoveva Vargas-Solar

Chargée de Recherche, CNRS, LIG, genoveva.vargas@imag.fr

<http://cedric.cnam.fr/summer-school/MDD2018/>

Centre Paul Langevin



Scientific situation

- Biannual thematic school since 2010
- Supported by the data bases community in France, joint to the conference BDA : Principles and applications of data management

The MDD School aims at:

- Present fundamental topics for understanding the evolution of research in the database domain
- Prepare human resources in the innovation process to come and trigger their evolution

Scientific situation

- **2010 Session:** Distributed systems for managing data and the semantic Web
- **2012 Session:**
 - Operators for Peta-octet scalable query processing
 - New information sources (social networks)
 - Data integration issues for independent environments
- **2014 Session :** Analytical visualisation of data, data quality and crowdsourcing
- **2016 Session :** Privacy respectful Big Data analytics processes

2018 Scientific Situation

Scalable Data Analytics

Data **revolution** encouraged by the **increasing storage & computing capacity** and its democratisation

- Increasing data collections availability their massive statistical analysis enables to progressively extract knowledge and make useful decision making for society
- The cost of data analysis, maintenance and security continues to grow because of
 - Few expertise in data analytics with formal education on the topic
 - The difficulty to organise massive, heterogeneous data processing processes where data stem from different sources
 - Few security ensured by software devoted to process and exploit data



Big Data exploitation needs several skills:

Know how to manage such data collections

Know how to analyse using statistical methods

School objectives, audience & impact



Specialized Human Resources Education

- Let (young) scientists better understand how to consolidate the position of their research and thereby reinforce the identity of their community
- Young scientists education: forum letting (young) scientists discuss their current work with colleagues and senior scientists for sharing ideas



Audience

- Young scientists (PhD students & postdocs) working in the database domain and related domains
- Lecturers, associate & senior professors and scientists for exploring new trends and creating a collaboration network or research reorientation possibilities
- Research engineers from the private sector.



This action will have
consequences...

Organization of the community :

- Create collaboration networks
- Maintain & encourage the cohesion within the DB community
- Open up research towards new topics and scientific communities

Education & dissemination :

- Transference of methods & techniques
- Feedback by senior scientists towards young researchers work

Program

Timetable	Participants	Module
Monday, June 18, 2018		
09:00-12:00	Michalis Vazirgiannis LIX, École Polytechnique	Graph based text Mining and the case of the Greek Web Archiving
14:00-17:00	Helena Galhardas Instituto Superior Técnico (IST), Lisbon	Data Quality Assessment and Improvement
Tuesday, June 19, 2018		
09:00-10:30	Vassilis Christophides Inria Paris	IoT analytics
11:00-12:30	Khalid Belhajjame University Paris Dauphine	Computational Reproducibility: Workflows, Provenance and Scripts
14:00-17:00	Nicolas Travers CNAM-Paris	Conducting Experiments (TP)
Wednesday, June 20, 2018		
09:00-12:00	Laurent d'Orazio University of Rennes	Big data management and a practical experience with TPC-H
14:00-17:00	Social event	TO BE ANNOUNCED
Thursday, June 21, 2018		
08:30-12:30	Vincent Leroy University of Grenoble Alpes	Data analysis on Apache Spark With a break at 10:00
14:00-17:00	Mustapha Lebbah University of Paris 13	Scalable Machine learning



<http://cedric.cnam.fr/summer-school/MDD2018/custom/programme>

Surf data in the French Alps !

Masses de données 17th
distribuées 22nd
Scalable Data Analytics **June**
2018
AUSSOIS
FRANCE



<https://www.facebook.com/events/2000992150163263/>



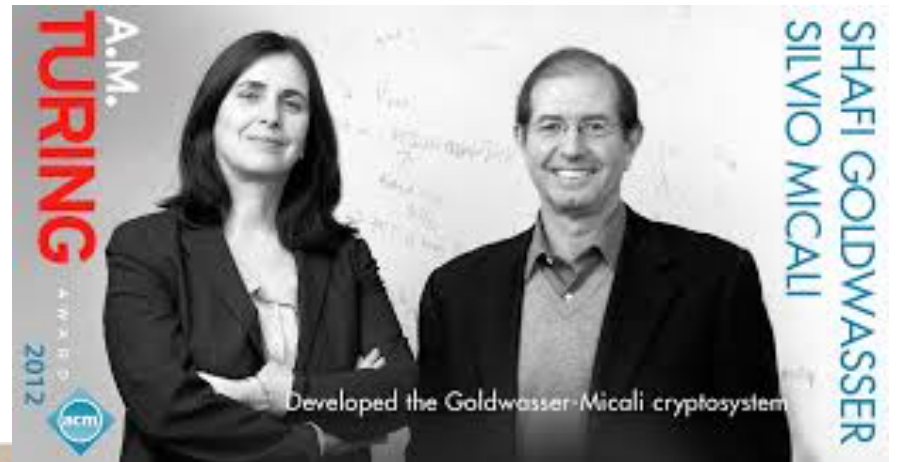
@BdaMdd

FRANCES ALLEN

First woman to receive the ACM Turing Award



A.M. TURING AWARD 2006



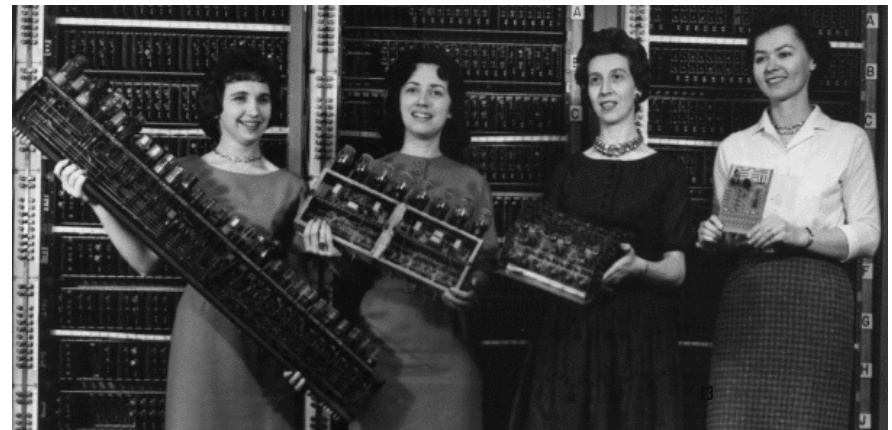
TURING

2008



BARBARA LISKOV

Developed the Liskov substitution principle



Genoveva Vargas-Solar

Chargée de Recherche, CNRS, LIG, genoveva.vargas@imag.fr

Cédric du Mouza

Maître de Conférences, CNAM, dumouza@cnam.fr

Luc Bouganim

Directeur de Recherche, INRIA, luc.bouganim@inria.fr

Fabrice Jouanot

MC, UGA, fabrice.jouanot@imag.fr

Javier A. Espinosa Oviedo

Postdoctorant, LIG, javiera.Espinosa@imag.fr

Nicolas Travers

MC, CNAM, nicolas.travers@cnam.fr

Christine Guimet

Pôle Budget-Contrats de recherche, LIG, Christine.Guimet@imag.fr

Samantha Sánchez

Assistante HADAS, LIG, samantha.sanchez@imag.fr

<http://cedric.cnam.fr/summer-school/MDD2018/>