

## Ph.D. Scholarship Application – 2018

### **Title :** A recommendation system for BigData analytic approaches.

Laboratoire LAMSADE – Université Paris Dauphine

#### **Offer:**

Funding doctoral contract

Starting date: September 2018.

Fixed term contract of 3 years

#### **Summary:**

Nowadays organizations look for Big Data as an opportunity to manage and explore their data with the objective to support decisions within its different operational areas. Indeed, the Big Data phenomenon has revolutionized and impacted the modern computing industry, which have reviewed their policies, architectures, and their production environment to support a continuous increase on the computational power that produces an overwhelming flow of data [3]. Big Data databases have recently become important NoSQL (being non-relational, distributed, open-source, and horizontally scalable) and NewSQL (taking the advantages of relational and NoSQL systems) data repositories in enterprises as the center for data analytics, while enterprise data warehouses (EDWs) continue to support critical business analytics. This scenario induced a paradigm shift in the large scale data processing and data mining, computing architecture, and data analysis mechanisms. This new paradigm has spurred the development of novel solutions from both industry (e.g., analysis of web-data, clickstream, network-monitoring logs) and science (e.g., analysis of data produced by massive-scale simulations, sensor deployments, telescopes, particle accelerators, genome sequencers) [4, 5].

In this context, analytical data management applications, affected by the explosion of the amount of generated data, are shifting away their analytical databases towards a vast landscape of architectural solutions combining storage techniques, programming models, languages, and tools [1]. These aspects represent the main features that allow distinguishing classical EDWs from analytical Big Data approaches. Besides, with massive growth of data generated from social network, connected sensors, mobile devices, recent analytical approaches addressed the paradigm of real-time analytic of streaming data [2].

In this scenario, non-expert users who have to decide which analytical solution is the most appropriate for their particular constraints and specific requirements in a Big Data context, is today lost, faced with panoply of disparate and diverse solutions.

**Objectives:** The aim of this thesis is to develop solutions to help (non-expert) users in the hard task of selecting the appropriate Big Data approach according their specific requirements.

#### **Required skills:**

The thesis will be mainly theoretical work, although specific cases will be used to test the theories.

We seek a student with:

- Good knowledge in Computer Science and Mathematics.
- Good knowledge of NoSQL, NewSQL, graphs Databases.
- Basic knowledge of machine Learning (desirable but not mandatory)
- Knowledge of analytic approaches,
- Good programming skills,
- Good command of written and spoken English.
- Master degree in Computer Science or equivalent degree giving access to PhD studies.

**Keywords** : Recommendation system, Analytic approaches, BigData.

**To apply:**

Please send the following material before April 15th, 2018 to [marta.rukoz@dauphine.fr](mailto:marta.rukoz@dauphine.fr) and [sonia.guehis@dauphine.fr](mailto:sonia.guehis@dauphine.fr):

- fully detailed CV,
- academic records (Master degree or equivalent),
- cover letter,
- recommendation(s) and supporting letter(s).

**Bibliography:**

[1] Yudith Cardinale, Sonia Guehis and Marta Rukoz: 'Big Data Analytic Approaches Classification'. ICSOFT, 12<sup>th</sup> international conference on Software Technologies, July 2017.

[2] Yudith Cardinale, Sonia Guehis and Marta Rukoz: 'Classifying Big Data Analytic Approaches: A Generic Architecture'. In series "Communications in Computer and Information Science" (CCIS) published by Springer ( in revision).

[3] Kune, R., Konugurthi, P.K., Agarwal, A., Chillarige, R.R., Buyya, R.: The anatomy of big data computing. *Softw. Pract. Exper.* 46(1) (2016) 79–105

[4] Chen, M., Mao, S., Liu, Y.: Big data: A survey. *Mob. Netw. Appl.* **19**(2) (2014) 171–209

[5] Philip Chen, C., Zhang, C.Y.: Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences* **275**(Complete) (2014) 314–347