

**3-year fully funded PhD position in the research area of Big Data and Artificial Intelligence (AI) in collaboration of the MIDI team of the ETIS Lab UMR 8051, Paris, France and the Department of Computer Science, University of Warwick under the supervision of Professor Dimitris Kotzinos and Professor Peter Triantafillou respectively.
(PhD funded under the Paris Seine Initiative of Excellence)**

Scope and Context

Due to the massive amounts of available data, various critical database tasks, e.g. query answering, become more of an approximate task than an exact one. On the other hand, the functioning of many critical Big Data system components depends on monitoring and predictions: e.g. in caching subsystems (which items to cache/prefetch), in query optimisation (the best access method to use), in indexing (when and for which attributes to build indexes). Additionally, big data analytics' systems need to be able to decide on the fly the most suitable (e.g. matching or optimization) algorithms to use in different cases. Similarly, many different prediction models for analytical queries (e.g., regression models) may perform differently for different predictive analytics tasks, so the system must decide on the best model to use. These problems can be approached by the use of predictive modelling adaptation techniques, well established in Artificial Intelligence (AI) and Machine Learning (ML). So, we propose to focus on working towards extending current Big Data management and analysis systems with ML and AI-based:

- Approximate analytical query processing engines based on ML models - e.g., queries based on descriptive statistics (COUNT, AVG, SUM, etc.) or on dependence statistics (CORR, CoVar, regressions, etc.). Given the massiveness of the current datasets, approximate query answering is one of the solutions we can employ in order to get responses in reasonable time and provide at the same time error feedback and control. At the same time, we want to introduce into the system uncertainty models with guarantees of maximum error and an understanding of the trade-off error vs time/costs during query processing.
- Self-learning capabilities, big data management and analysis systems should be able to learn by monitoring operations and decisions made so far and use them to extract useful information in order to optimize various of the system's operation, like selecting the best possible algorithms, models, etc.

So, during this PhD we want to investigate the above issues and develop solutions that can be integrated to real world big data management systems.

We expect the successful applicant to be one of the driving forces behind the newly established collaboration between the two entities mentioned above. The successful applicant will work jointly with Professor Dimitris Kotzinos (ETIS / Paris Seine University) and Professor Peter Triantafillou (Department of Computer Science, University of Warwick) and their respective groups, will be based at ETIS lab at the University of Cergy Pontoise in the greater Paris area but frequent exchanges and stays at Warwick are envisioned.

Tentative Starting Date: October 2018

Application

If interested, please send your application (including a detailed CV, university transcripts, a copy of the master thesis and/or scientific papers if available, as well as a list of personal references and a motivation letter) in PDF format to Professor Dimitris Kotzinos (Dimitrios.Kotzinos@u-cergy.fr) and Professor Peter Triantafillou (P.Triantafillou@warwick.ac.uk). Further enquires are also welcome. Applications are welcome until 20/08/2018 or until the position is filled.