Big Data Algorithms and Programming in Spark and Hive.

Mohamed-amine Baazizi*[1] and Dario Colazzo*†[2]

[1]Laboratoire dÍnformatique de Paris 6 (LIP6) – Université Pierre et Marie Curie - Paris 6, Centre National de la Recherche Scientifique : UMR7606 – 4 Place JUSSIEU 75252 PARIS CEDEX 05, France
[2]Laboratoire dánalyse et modélisation de systèmes pour láide à la décision (LAMSADE) – Université Paris-Dauphine, Centre National de la Recherche Scientifique : UMR7024 – Place de Lattre de Tassigny 75775 PARIS CEDEX 16, France

Résumé

The tutorial is about algorithms and programming on the map-reduce based framework Spark and Hive.
After a brief recall of the map-reduce paradigm, attendees will have the possibility of getting familiar with Spark main mechanisms by designing and implementing algorithms for text precessing, matrix manipulation and graph analytics. In a second part, attendees will be introduced to Hive and will design algorithms for large-scale Web-log analytics.

Mohamed-Amine Baazizi

holds an Associate Professor position at Université Pierre et Marie Curie since September 2013. Prior to that, he was a postdoctoral researcher in Telecom Paris tech. He obtained his PhD from University of Paris Sud in September 2012. His research focuses on the management of large scale semi-structured data like JSON and XML. He is recently investigating topics related to the representation and flexible querying of incomplete data.

Dario Colazzo

Dario Colazzo is Full Professor at Université Paris Dauphine. His research interests focus on databases and programming languages, and include cloud databases and type systems for safe and efficient processing of semi-structured data. Past research interests included type systems for the lambda-calculus and pi-calculus.

*Intervenant
†Auteur correspondant: dario.colazzo@lamsade.dauphine.fr