

# Big Data Algorithms and Programming in Spark and Spark SQL.

**Mohamed-Amine Baazizi**

mohamed-amine.baazizi@lip6.fr

Laboratoire d'Informatique de Paris 6 (LIP6)

Sorbonne Université

Centre National de la Recherche Scientifique :

UMR7606, 4 Place JUSSIEU 75252 PARIS CEDEX

05, France

**Dario Colazzo**

dario.colazzo@lamsade.dauphine.fr

Laboratoire d'Analyse et Modélisation de

Systèmes pour l'Aide à la Décision (LAMSADE)

Université Paris-Dauphine, Centre National de la

Recherche Scientifique : UMR7024

Place de Lattre de Tassigny 75775 PARIS CEDEX

16, France

The tutorial is about algorithms and programming on the map-reduce based framework Spark. After a brief recall of the map-reduce paradigm, attendees will have the possibility of getting familiar with Spark's main mechanisms by designing and implementing algorithms for text processing, matrix manipulation and graph analytics. In a second part, attendees will be introduced to Spark SQL and will query structured and some tricky semi-structured data.

**Mohamed-Amine Baazizi.** Mohamed-Amine Baazizi holds an Associate Professor position at Sorbonne Université 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** is full Professor in Computer Science at University Paris Dauphine, PSL, in the LAMSADE laboratory. Since 2016 he is adjunct Professor (part-time) at École Polytechnique. His research and teaching activities are in the context of formal systems and algorithms for efficient processing of massive semi-structured datasets (JSON, RDF and property graphs), mainly relying on shared-nothing parallelism and distribution (Hadoop, Spark, Flink, ...). He leads the Data Science team at LAMSADE, and he is director of the Executive Master in Artificial Intelligence and Data Science, at the DEP department. He is regularly invited to participate to program committees of international conferences, such as EDBT, SIGMOD, IJCAI, WISE, VLDB, in recent years.