Un environnement à grande échelle pour le traitement de flots massifs de données

Mots clés :
- Directeur de thèse : Pierre Sens
- Co-encadrant(s) :
- Unité de recherche : Laboratoire d'informatique de Paris 6
- Ecole doctorale : École Doctorale Informatique, Télécommunications, Électronique de Paris
- Domaine scientifique principal : Divers

Résumé du projet de recherche (Langue 1)
Managing and processing Big Data is usually handled in a static manner. Static cluster- or grid-based solutions are ill-suited for tackling Dynamic Big Data workflows, where new data is produced continuously. Our final objective in this thesis is to design and implement a large scale distributed framework for the management and processing of Dynamic Big Data.

Résumé du projet de recherche (Langue 2)
We will especially focus our research on the following scientific challenges: 1. Dynamic Big Data Placement. One of the most challenging problems is to place new data coming from a huge workflow. We expect to propose new techniques for the mapping of huge dynamic flows of data in a large scale distributed system. In particular, these techniques ought to promote the locality of distributed computations. 2. Dynamic Big Data Processing. Our objective here is to propose innovative solutions for the processing of a continuous flow of big data in a large scale distributed system. We expect to propose new techniques to support such computations. We will identify properties that are common to distributed programming paradigms. We will then integrate the properties in the design of a framework that takes into account the locality of the data flow, and ensures a reliable convergence of the data processing.