



CONFÉRENCE DÉBAT

« La Datamasse : directions et enjeux pour les données massives »

Organisée par Serge ABITEBOUL et
Patrick FLANDRIN, Membres de l'Académie des sciences

Mardi 18 Février 2014 de 14h30 à 17h30
Académie des sciences - Grande salle des séances
23 quai de Conti, 75006 Paris

Résumé : Nous vivons dans une « société de l'information » dont les avancées scientifiques et techniques rapides, associées au développement d'usages nouveaux, conduisent à produire des quantités toujours plus gigantesques de données numériques. Cette situation d'abondance ouvre des perspectives nouvelles tant dans les sciences exactes que dans les sciences humaines. L'utilisation de cette « datamasse » (Big Data en anglais) pose des défis considérables : Comment stocker de telles quantités de données, les manipuler, les analyser, les trier, ... les valoriser ? Comment concilier leur omniprésence et le respect de la vie privée ? Comment faire qu'elles bénéficient à tous ? Ce sont quelques-uns de ces aspects qui seront mis en avant dans cette rencontre, afin d'en mieux comprendre les possibilités et les limitations, pour en mieux maîtriser les développements.

14h30

Introduction

Serge Abiteboul, Directeur de recherche Inria, École Normale Supérieure de Cachan, Membre de l'Académie des sciences et **Patrick Flandrin**, Directeur de recherche CNRS, École Normale Supérieure de Lyon, Membre de l'Académie des sciences

14h45

À la découverte des connaissances massives de la Toile

Serge Abiteboul, Directeur de recherche Inria, École Normale Supérieure de Cachan, Membre de l'Académie des sciences

15h15

Des mathématiques pour l'analyse de données massives

Stéphane Mallat, Professeur à l'École Normale Supérieure, Paris

15h45

La découverte du cerveau grâce à l'exploration de données massives

Anastasia Ailamaki, Professeure à l'École Polytechnique Fédérale de Lausanne

16h15

Big Data et Relation Client : quel impact sur les industries et activités de services traditionnelles ?

François Bourdoncle, co-fondateur et CTO d'Exalead, filiale de Dassault Systèmes

16h45

Discussion générale & Conclusion



À la découverte des connaissances massives de la Toile

Serge Abiteboul, *Directeur de recherche Inria, École Normale Supérieure de Cachan,
Membre de l'Académie des sciences*

La Toile nous propose des masses d'information et de données considérables. Avec les réseaux sociaux, ses utilisateurs communiquent, publient, interagissent, développant de nouveaux modes de relations, et de pensée. Nous parlerons de challenges pour trouver les connaissances sur la Toile, les évaluer, valider, vérifier, classer...

Des mathématiques pour l'analyse de données massives

Stéphane Mallat, *Professeur à l'École Normale Supérieure, Paris*

L'accumulation massive de données permet d'analyser des phénomènes complexes qui dépendent d'un très grand nombre de variables inconnues. Les applications sont considérables, pour la reconnaissance de signaux, les diagnostics médicaux ou industriels, ou le traitement de données non structurées, provenant notamment de l'Internet. Stocker des informations n'est cependant pas suffisant pour affronter la malédiction de la dimensionnalité, qui explose avec le nombre de variables. Les développements algorithmiques récents ouvrent des problèmes mathématiques profonds qui touchent la géométrie en grande dimension, l'analyse harmonique, les EDP, et bien sûr les probabilités et statistiques. Le cerveau ayant des performances de traitement extraordinaires, ces questions sont aussi en interface avec les neurosciences, comme on le verra.

La découverte du cerveau grâce à l'exploration de données massives

Anastasia Ailamaki, *Professeure à l'École Polytechnique Fédérale de Lausanne*

Les processus scientifiques dépendent aujourd'hui fortement de l'analyse rapide et précise des données expérimentales. Les scientifiques sont régulièrement dépassés par l'effort nécessaire pour gérer les volumes de données produites par l'observation de phénomènes, ou par des simulations sophistiquées. Les logiciels de gestion des données se révélant inefficaces, inadéquats ou insuffisants pour répondre aux besoins des applications scientifiques, la communauté scientifique utilise généralement des logiciels ad hoc spécialisés. Cependant, avec la croissance exponentielle du volume des données et avec la complexité de celles-ci, ces systèmes spécialisés ne passent pas à l'échelle et ne permettent plus d'analyser efficacement les données, ce qui ralentit le cycle d'analyse, la compréhension et la préparation de nouvelles expériences. Je vais illustrer le problème avec une application complexe sur des données de simulation du cerveau ; je vais montrer comment les problèmes de neurosciences se traduisent par des défis pour la communauté de gestion des données. Je vais montrer comment une nouvelle technologie de gestion des données peut permettre aux neuroscientifiques de simuler et de découvrir une part significative du cerveau humain à des niveaux de détail sans précédent. Enfin, je vais décrire les défis de l'intégration des données de simulation et des données médicales de neurosciences pour faire avancer notre compréhension du fonctionnement du cerveau.



Big Data et Relation Client : quel impact sur les industries et activités de services traditionnelles ?

François Bourdoncle, *co-fondateur et CTO d'Exalead, filiale de Dassault Systèmes*

Le Big Data est un mot à la mode dont beaucoup se demandent ce qu'il recouvre. Au sens littéral « analyse des données massives », le Big Data est l'ensemble des techniques et des applications nouvelles permettant la valorisation des données gigantesques récoltées par toutes nos activités, personnelles ou professionnelles, à l'ère numérique. Mais au-delà des enjeux technologiques, réels, des techniques hybrides à la frontière des statistiques, de l'apprentissage automatique et de l'algorithmique distribuée, se cachent des enjeux d'une ampleur toute autre. Cet exposé va tenter de mettre en lumière la manière dont le Big Data est utilisé par les puissants acteurs de l'économie numérique pour remettre en question la relation client existante dans nombre d'industries et d'activités de service, afin de ré-intermédiaire ces dernières et de capter ainsi une partie de plus en plus importante de la chaîne de valeur.