

Simplexité des Big Data : la simplicité au service de la complexité

par François STEPHAN (91)

Le nouveau paradigme du Big Data est au cœur des nouveaux enjeux de la recherche technologique française pour les systèmes du futur de plus en plus complexes.

Un Buzz word qui rappelle celui du Cloud

« Big Data » a été élu mot numérique de l'année 2012, succédant ainsi à « Cloud Computing » élu en 2011¹; personne ne s'en étonnera. S'agit-il du nouveau « Buzz word » du monde de l'informatique qui en est friand ou bien un phénomène bien réel annonçant de grandes transformations ? Disons-le ici clairement, nous sommes convaincus de nous situer au-delà du Buzz word, et de vivre avec le Big Data une révolution majeure de notre économie au sens large, appuyée par des évolutions techniques, elles-mêmes alimentées par les progrès constants de la micro-électronique et la loi de Moore.

Big Data = Smart Data

Lors des Entretiens Telecom Paristech sur le Big Data en décembre 2012, Pierre Barnabé (SFR) a défini le Big Data comme le moyen de « rendre intelligents et intelligibles des flux de données exponentiels ». Autrement dit, Big Data = Smart Data.

Le projet Google Flu (Suivi de la grippe) est un bel exemple d'application du Big Data aux moteurs de recherche. Les



chercheurs de Google ont identifié une combinaison de 45 termes utilisés dans les recherches effectuées dans leur moteur de recherche, semblant être de bons indicateurs de la propagation de la grippe, qui se corrélaient très bien (via une modélisation mathématique), pour donner des prédictions fiables sur la couverture géographique de la grippe saisonnière extrêmement semblables avec les statistiques officielles issues des consultations médicales sur le terrain². On imagine facilement ici l'intérêt d'un tel outil, rendu disponible gratuitement par Google.

La simplicité du Big Data : un nouvel enjeu majeur pour la recherche française

«La Simplicité est l'art tout autant que la science de trouver et activer des solutions simples, efficaces et élégantes à des enjeux décisifs complexes.»³

Est-ce que le nouveau paradigme du Big Data peut répondre en partie aux enjeux de l'ingénierie numérique de systèmes de plus en plus complexes ?

Les systèmes complexes, que sont par exemple les systèmes de transport multimodaux, rassemblant systèmes ferroviaires urbains, systèmes d'auto-partage, autobus, etc., demandent de nouvelles modélisations qui prennent en compte des données jusqu'à présent absentes du périmètre de jeu, par exemple les données produites en temps réel sur le déplacement des usagers. De telles données, associées aux données disponibles en mode « Open data », appellent les approches méthodologiques et technologiques du Big Data.

L'IRT SystemX (Institut de Recherche Technologique dédié à l'Ingénierie des systèmes du futur et soutenu par le Plan d'Investissement d'Avenir), porté par le campus Paris Saclay et labellisé par le pôle de compétitivité «Systematic Paris-

Region», comporte deux programmes de recherche appliquée : « Systèmes de systèmes » et « Technologies & outils d'ingénierie numérique », ainsi qu'un programme de formation et de relations PME.

Les projets de recherche du programme « Systèmes de Systèmes » rassemblent sur un même site, dans un mode « intégratif », des industriels (grands groupes, PME et ETI) de secteurs différents partageant des enjeux Systèmes proches, sur les trois thématiques que sont le Transport multimodal, la Sécurité, et la Gestion de l'Energie, en les associant avec les laboratoires académiques les plus pertinents, et des chercheurs en propre de l'IRT.

Les projets de recherche du programme « Technologies et Outils » se concentrent, eux, sur la maîtrise de systèmes et logiciels embarqués, de moyens de calcul à hautes performances, d'infrastructures de communication, de production et de distribution de l'énergie numérique (ou Cloud Computing), et la maîtrise de la modélisation, de la conception, de l'intégration et de la validation de grands systèmes, ainsi que sur le contrôle des outils d'ingénierie numérique pour le développement de maquettes numériques multi-échelles et multi-physiques.

Ainsi, l'IRT SystemX mène des projets de recherche rassemblant grands industriels, PME innovantes et laboratoires académiques, visant à résoudre les défis technologiques et scientifiques posés par l'ingénierie numérique de Systèmes complexes, au sein desquels le Big Data représente de plus en plus un enjeu scientifique tout comme une solution technologique. On pourra citer par exemple l'utilisation de techniques de DataViz pour une plateforme de simulation permettant d'analyser des résultats issus de niveaux de description différents et de nature hétérogène, ou encore la mise à disposition des architectures Hadoop « à la demande » par les opérateurs de Cloud Computing.

Citons également un projet de recherche appliquée que mène l'IRT SystemX sur

les technologies visant à extraire des informations significatives à partir de flux très importants de données provenant de médias et de langues multiples, en mettant en œuvre certaines des technologies du Big Data associées à la science de la linguistique et au Web sémantique. Extraire ainsi des signaux faibles parmi des données fortement bruitées : l'une des applications du Big Data pour l'analyse d'opinion au travers des réseaux sociaux, par exemple.

Dans le même registre, la numérisation et le passage à la génération « intelligente » des réseaux électriques (les « smart grids »), associés aux architectures de type Internet & Social, sont de nature à nécessiter les approches Big Data de manière très directe. Les industriels qui participent aux projets de recherche de l'IRT SystemX au sein de sa thématique « Gestion intelligente de l'Energie » ont positionné le Big Data parmi les défis technologiques et scientifiques majeurs à résoudre.

L'excellence scientifique française au service de l'industrie

L'excellence française des laboratoires de recherche académique en mathématiques (algorithmes notamment), statistique, sémantique, linguistique, design (compétence clé pour le DataViz), et plus largement en sciences humaines, sont des atouts majeurs que l'IRT SystemX met au service des industriels - grands groupes, PME et ETI - qui participent à ses projets de recherche à contenu Big Data plus ou moins fort.

Autre signe de la réalité du nouveau paradigme du Big Data, le pôle de compétitivité Systematic-Paris-Region, qui fédère en Ile de France plus de 700 acteurs industriels, PME et scientifiques, introduit pour la première fois le Big Data comme l'un des six axes technologiques prioritaires de son nouveau plan stratégique 2013-2018, parmi les plus porteurs de transformations et d'innovations.

Les thèmes principaux de recherche identifiés par le pôle Systematic sur cette thématique portent notamment sur l'Analyse et la fouille de données, la Modélisation par les données, et les Outils et méthodes de représentation et visualisation (DataViz).

Big Data, évolution technologique, révolution des usages... et révolution économique et sociétale ?

Tout comme Le Cloud computing, le Big Data, évolution technologique, est une véritable révolution des usages qui implique une révolution managériale au sein des acteurs économiques, privés et publics. Après l'entreprise Cloud, voici venir l'entreprise Big Data, l'entreprise numérique, digitale, au sein de laquelle "la ressource première est bien l'information ou en tout cas la transformation du bruit en information pertinente"⁴.

L'ère du cognitif n'est plus très loin, et avec elle la promesse d'apporter la simplicité nécessaire pour le management de Systèmes de plus en plus complexes, et forcément de plus en plus numériques, au service de la société du 21^{ème} siècle. ■

L'AUTEUR



François STEPHAN est Directeur du Programme "Systèmes de Systèmes" au sein de l'Institut de Recherche Technologique SystemX, dédié à l'ingénierie

numérique des systèmes. Diplômé de l'Ecole polytechnique et de Telecom ParisTech, il a plus de 20 d'expérience professionnelle dans les technologies de l'information.

Il est le co-auteur avec Jean-Pierre BRIFFAUT de l'ouvrage "Cloud computing : Evolution technologique, révolution des usages" publié en mai 2013 par les éditions Hermes Science - Lavoisier.

¹ Election du mot numérique 2012 de David Fayon <http://davidfayon.fr/2013/01/big-data-mot-numerique-2012/>

² <http://www.google.org/flutrends/intl/fr/about/how.html>

³ Grégory Le Roy, 25/11/2011, Le Cercle Les Échos

⁴ Bruno Teboul, 28/02/2013, Le Cercle Les Échos