



organisent en partenariat avec



www.uimm.com



www.captronic.fr/ www.telecom-paristech.fr/nc/formation-et-innovation-dans-le-numerique.html

Le potentiel et les défis du Big Data

UIMM

Mardi 2 et mercredi 3 juillet 2013

56 avenue de Wagram 75017 PARIS



Source : belogik.com

Le potentiel et les défis du Big Data

Mardi 2 et mercredi 3 juillet 2013

UIMM, 56 avenue de Wagram, 75017 PARIS

Les données s'abattent à rythme exponentiel sur les organisations. Les nouveaux systèmes d'informations, les vecteurs d'expression, les systèmes de paiement, et l'ouverture de multiples bases de données publiques et privées génèrent chaque jour des afflux soudains de milliards d'informations. De plus, chaque acteur (particulier, administration, organisation, entreprise, groupement communautaire) se fait directement producteur de nouveaux corpus d'informations non ou semi-structurées : données personnelles, applications géolocalisées (conjuguées à la dimension temporelle), conversations sur réseaux sociaux, événements, contenus dématérialisés, photos Flickr, microblogs... D'autre part, les communications issues de la multiplicité des objets communicants numériques (internet des objets, communication M2M, capteurs, sondes, RFID) génèrent de la donnée à grande échelle.

Le Big Data consiste à intégrer, synchroniser, traiter et valoriser ces flux de données vertigineux, extrêmement variés et de différentes natures. Il s'agit à terme de donner du sens à des gisements massifs de données et d'en faire ressortir la quintessence et la valeur ajoutée. Les data ne sont plus de formats dits structurés et relationnels, mais référencées comme contenus non structurés et hétérogènes (commentaires, micro-discussion, vidéos, images, sons, données de capteurs ...). Aller collecter un document posté relève d'une donnée structurée. En revanche, en comprendre le contenu, faire de la reconnaissance d'images ou d'objets, définir les variables explicatives, résoudre les subtilités sémantiques, croiser différentes sources de données relèvent d'une analyse non structurée. La nature de ces données hétérogènes ne connaît aucune limite pour le Big Data.

Pour ce séminaire, nous avons demandé à quelques-uns des meilleurs experts français de grands groupes, mais aussi de jeunes entreprises innovantes, de présenter les derniers développements en matière de Big Data. La première journée après quelques présentations d'introduction, sera consacrée à l'état de l'art et des technologies du Big Data. La seconde journée abordera les applications qui mettent en valeur l'utilisation du Big Data.

Au sommaire du Mardi 2 juillet 2013

ENJEUX ET PERSPECTIVES DU BIG DATA

9h – 9h30 : Vulgarisation et décryptage de l’impact du Big Data

Par Matthias FILLE, Conseiller Développement International filière TIC, CCI Paris Il- de - France

Lors de cette intervention, nous reviendrons brièvement sur l’émergence et la genèse du Big Data qui représente une rupture de paradigme dans l’analyse des données. Le propos suivant se voudra volontairement illustratif et empirique. Nous évoquerons quelques exemples d’applications business, philanthropes ou encore transposés dans la sphère médicale.

9h30 -10h : Le marché du Big Data : évolutions technologiques et business

Par Olivier RAFAL, Directeur de recherches Logiciels - Information Management Secteur public chez PAC

Au-delà de l’habillage marketing de procédés et techniques souvent déjà connus et mis en oeuvre, le Big Data caractérise une évolution majeure de la sphère de l’aide à la décision. Grâce aux nouvelles technologies, l’accès à ces techniques s’est en effet démocratisé et les entreprises ont devant elles un vaste champ d’expérimentation, qu’elles peuvent exploiter pour optimiser leurs décisions commerciales, marketing, stratégiques... Cette promesse du Big Data se traduit, selon le cabinet d’étude et de conseil Pierre Audoin Consultants, par une grande effervescence sur le marché ; la croissance annuelle moyenne des logiciels et des services en France sur ce segment devrait dépasser les 40% d’ici à 2016!

Cette session reviendra sur les grandes évolutions technologiques et leurs conséquences sur les acteurs et sur le marché."

10h – 10h30 : Aspects juridiques : quelles données pour quels usages ?

Par Olivier DE COURCEL, Avocat associé chez FERAL-SCHUHL / SAINTE-MARIE

Le phénomène du « Big Data » traduit un changement de dimension et peut-être même de nature dans les données traitées par nos entreprises.

Mais de quelles données s'agit-il : publiques, privées, personnelles, anonymes, autres ? font-elles seulement l'objet de droits ? Et à qui appartiennent-elles : à la personne qu'elles concernent ? A l'entreprise qui les collecte ou qui les propose ? A quelqu'un d'autre ou à personne ? Quel prix peut-on y attacher : un prix de collecte, un prix de traitement, ou bien un prix d'usage ? Une entreprise peut-elle restreindre l'utilisation ou la réutilisation de « ses » données ?

Ces questions doivent être vues non seulement au regard du droit et de la jurisprudence actuels, mais aussi à la lumière des évolutions annoncées, notamment le projet de règlement européen sur les données à caractère personnel. Elles sont aussi éclairées par les initiatives étrangères, notamment l'« Open Data » et son impact sur la diffusion de données par les pouvoirs publics. L'approche juridique apporte enfin des réflexions sur la maîtrise ou la gestion des données par les individus (« Small Data »).

Ce sont ces différents aspects du « Big Data » que cette présentation permettra de voir ou de revoir.

10h30 – 11h00 : Pause Café

ETAT DE L'ART ET TECHNOLOGIES DU BIG DATA

11h00-11h40 : Entreprise et Big Data

Par Christophe FAVART, Chief Development Architect chez SAP

La présentation s'attachera à décrire quels sont les challenges auxquels font face les entreprises à ce jour en terme de gestion des données, que représentent les différentes étapes « Big Data », considérant les données internes et externes et comment valoriser cette masse de données en terme de découverte et d'analyse.

11h40 – 12h20 : Aperçu de l'écosystème Hadoop et de son utilisation en France

Par Nicolas THIEBAUD, fondateur du Hadoop User Group en France (HUG France) et directeur technique de CAPTAIN DASH

Nicolas THIEBAUD fera la présentation des technologies Hadoop et des cas d'utilisations observés en allant de Map Reduce jusqu'à Impala. Nous concluons par le retour d'expérience Captain Dash.

12h20 – 13h : L'analyse en temps réel des informations dans un contexte BigData

Par Jérôme CHAILLOUX, Europe IOT - Sr. Technical Field Specialist - Big Data, Linux Advocate IBM Software

IBM InfoSphere Streams permet une analyse continue et extrêmement rapide de volumes massifs d'informations en mouvement afin de faciliter la connaissance de l'activité métier et le processus décisionnel. Système de traitement ultra-performant qui analyse rapidement l'information diffusée par plusieurs milliers de sources en temps réel, augmentant la rapidité et l'exactitude du processus décisionnel dans différents domaines tels que la santé, l'astronomie, le secteur manufacturier, les transactions financières, etc. InfoSphere Streams fournit une plateforme et des services d'exécution aux applications développées par les utilisateurs qui absorbent, filtrent, analysent et mettent en corrélation des volumes parfois énormes de flux de données continus. Il prend en charge la composition de nouvelles applications sous la forme de graphiques de traitement des flux pouvant être créés instantanément, mis en correspondance avec diverses configurations matérielles, et adaptés au fur et à mesure de la réception et de l'envoi des demandes, et des modifications des priorités relatives. InfoSphere Streams prend en charge de gros volumes de données transmises en continu, structurées et non structurées : images, son, voix, VoIP, vidéo, TV, actualités financières, radio, scanners policiers, trafic Web, email, chat, données GPS, données de transaction financières, données par satellite, capteurs, lecteurs de badges, etc.

13h -14h30: Déjeuner

14h30 – 15h20 : S'attaquer au Big Data avec Hadoop et l'intégration Open Source

Par Fabrice BONAN co-fondateur de TALEND et Chief Operating Officer

Les entreprises sont confrontées à des quantités croissantes de données et les problèmes liés aux demandes de volume et d'utilisation ont forcé les responsables informatiques et développeurs à rechercher de nouvelles solutions. Cela s'est traduit par de l'innovation dans les domaines du traitement massivement parallèle et du stockage de données non-relationnelles. Apache Hadoop, une plate-forme de logiciel open source, s'est imposé comme la

technologie de référence pour les grandes entreprises aux besoins d'analyses sophistiquées et de transformation de pétaoctets de données complexes.

Fabrice Bonan abordera l'accès, la manipulation et le stockage de grands volumes de données avec Hadoop et la possibilité de bénéficier de hautes performances, d'optimiser les coûts d'intégration de données avec une importante évolutivité.

Au cours de cette session, les participants apprendront à :

- tirer parti de la croissance des données ;
- mettre en production et exploiter l'architecture d'Hadoop ;
- traiter d'importants volumes de données complexes en combinant Hadoop et une architecture open source.

15h20 – 16h : Comment le Big Data change la business intelligence ? L'exemple du Search Engine Optimisation

Par Vincent HEUSCHLING, AFFINI-TECH SARL

Les technologies Big-Data et leur chef de file Hadoop révolutionnent l'informatique décisionnelle. Ces changements s'opèrent à plusieurs niveaux que sont tant l'**aspect économique** avec une forte démocratisation des solutions, que sur l'approche, avec de nouveaux paradigmes visant à **systematiser le Data-Mining et le Machine-Learning**. Par ailleurs on constate que ces technologies Big-Data peuvent apporter énormément à tout type d'organisations. Ce n'est pas seulement une question de volumes vertigineux à absorber, mais plus de **richesse du patrimoine informationnel**. L'**hybridation des données publiques** (Open-Data, réseaux sociaux) et privées peuvent apporter des dimensions complémentaires aux données internes des entreprises, et offrir un éclairage novateur aux managers.

Par l'illustration de cas clients simples vous percevrez comment l'utilisation des données et les prises de décisions ont pu évoluer. Ces cas clients nous permettrons de voir en détail les piles technologiques utilisées, et de mesurer la richesse de l'écosystème technique Big-Data Hadoop. C'est toute la logique de travail sur les catégories et les segments qui sont affectées par la capacité à travailler avec la granularité la plus fine imaginable. Cette approche permet de donner aux marketeurs et aux managers des éléments de décisions suffisamment

précis pour pouvoir envisager de réellement personnaliser la relation avec des clients.

16h – 16h30 : Pause café

16h30 – 17h00 : Big Data mais basse consommation

Par Laurent Julliard, KALRAY

De nombreux secteurs industriels comme l'automobile, le médical, la communication, produisent de plus en plus de données et réclament des méthodes de traitement toujours plus diversifiées et sophistiquées exigeant toujours plus de puissance de calcul mais surtout, et c'est nouveau, d'opérer à budget énergétique contraint. L'arrivée sur le marché des nouvelles architectures de processeurs dites manycore à la fois performantes et énergétiquement très efficaces ouvrent de nouveaux champs d'applications dans le traitement des données.

Durant cette intervention des exemples d'applications exigeantes en puissance de calcul et en consommation énergétique seront présentés ainsi que le processeur MPPA MANYCORE à 256 cœurs de la société KALRAY qui a permis leur mise en œuvre. Ce sera aussi l'occasion d'échanger sur les champs d'utilisation possibles de ces nouveaux processeurs dans le domaine du big data.

17h – 17h30 : Big Data: quelle formation permettant de répondre à la demande pressante de data scientist?"

Par Stéphane Cléménçon, professeur à TELECOM PARISTECH

A l'ère "Big Data", le concept d'une information « propriétaire », sous un format rigide standardisé, stockée dans une data-warehouse et circulant en circuit (très) fermé, semble être aujourd'hui totalement dépassé dans de nombreux domaines émergents. Les « Big Data » ne prenant leur sens/valeur qu'au sein d'un « écosystème » et ne pouvant être dissociées des usages/services qui sont réalisés à partir de ces dernières et/ou qui les génèrent. Le phénomène « Big Data » pose ainsi de nombreuses questions, relatives au développement technologique qu'il va devoir susciter (pour permettre en particulier l'exploitation effective des « big data »), à la valeur qu'il est susceptible de créer et renvoie à des problèmes scientifiques, sociétaux et juridiques, de nature différente mais souvent entremêlés. Il appelle aussi de nouveaux programmes de formation, intégrant ses

multiples facettes, les offres d'emploi relatif au métier de « data scientist » affichant une hausse spectaculaire sur le marché du travail ces dernières années. Une formation pluridisciplinaire, débouchant sur un savoir-faire opérationnel, serait particulièrement adaptée au domaine « Big Data », à l'interface de l'informatique et des mathématiques appliquées et appelant une vision d'ensemble stratégique pour la conception de nouveaux services/produits et le déploiement de systèmes d'information décisionnels avancés.

17h30 – 18h : MapR : La convergence du stockage et d'Hadoop pour l'entreprise

Par Aurélien GOUJET, Directeur Technique Europe du Sud et Benelux et Xavier GUERIN, VP Europe du Sud et Benelux

Au sein de l'écosystème Hadoop et des ses différentes distributions se distingue la société MapR, qui à travers une solution Hadoop 100% compatible open source, propose toutes les couches classiques d'Hadoop, mais enrichies par des fonctionnalités qui permettent la mise en production d'applications critiques d'entreprises avec un taux très élevé de disponibilité.

MapR offre fiabilité, utilisation aisée et une performance pour les entreprises du secteur financier, des médias, de la santé, de l'industrie, des telco et du gouvernement.

Au sommaire du Mercredi 3 juillet 2013

9h – 9h40 : Les technologies Big Data, un nouveau socle pour le SI d'entreprise?

Par Ghislain MAZARS, Président de UBEEKO SAS

Après avoir transformé l'architecture des centres de données des géants du web tels que Google, Yahoo et Facebook, les technologies Big Data font désormais partie intégrante des offres des principaux fournisseurs IT. Au-delà des premières preuves de concept, quel peut-être leur impact à long terme sur l'architecture des SIs d'entreprise? Après la virtualisation des applications, une nouvelle pile technologique va-t-elle progressivement s'imposer?

Après un rappel de l'histoire et des apports des technologies Big Data, la présentation s'attachera à montrer leur application à un SI d'entreprise à travers la solution Ubeeko de gestion et de gouvernance des données d'entreprise.

9h40 – 10h20 : Conserver les Big Data, source de valeur pour demain

Par Patrick FOSTER, CEO d'EXABUILDER

Les technologies d'analyse des données « Big Data » deviennent de plus en plus performantes, ce qui accroît le potentiel de valeur de ces données dans le futur.

De ce fait, il devient intéressant de conserver ces données sur le long terme pour les ré-analyser plus tard, le cycle de vie de ces données et de leur valeur pouvant atteindre plusieurs dizaines d'années.

Entre deux analyses, les données vont être totalement inactives, et leur coût de stockage peut être optimisé pour faire face à la montée à un rythme exponentiel de leur volume, en utilisant des moyens de stockage adaptés.

Ainsi il sera possible de ne pas se priver d'une source de valeur si précieuse dans le futur.

Après être revenu sur le cycle de vie des données « Big Data », nous reviendrons sur les grandes problématiques liées au stockage de ces données, et les moyens qu'apporte Exabuilder pour organiser leur conservation dans le temps et optimiser leur coût de stockage.

10h20 – 10h50 : Pause café

10h50 – 11h30 : Big Data : usages et perspectives

Par Mouloud DEY, Directeur Solutions et Marchés Emergents chez SAS

Les Big Data ont rapidement conquis l'espace médiatique, mais au-delà du thème classique de leur stockage, les aspects métiers en constituent le véritable enjeu. Dans cette nouvelle économie où les entreprises se doivent d'identifier de nouveaux relais de croissance, l'information (la donnée) constitue un actif essentiel au cœur de leurs stratégies de développement.

Si ces promesses de transformation, voire de « révolution industrielle » peuvent effrayer certaines organisations, elles constituent, pour les plus innovantes, des opportunités uniques en termes de compétitivité et de création de valeur.

Des expériences réussies dans des secteurs aussi variés que la banque (gestion du risque de liquidité), la grande distribution (optimisation de la marge) , les télécommunications (optimisation des campagnes marketing) , le secteur public (détection de fraude), montrent à quel point un usage différent et innovant des données accumulées dans les systèmes d'informations conduisent à des bénéfices opérationnels concrets susceptibles d'améliorer et de transformer sensiblement les modèles économiques de ces organisations.

11h30 – 12h20 : Retour d'expérience : Analyse prédictive en assurance Santé, grâce au Machine Learning et au NLP

Par Julien CABO, Directeur Big Data Analytics chez OCTO Technology

Les dépenses de santé sont soumises à un facteur psycho-social important qui pèse sur les remboursements des sociétés d'assurance et les mutuelles. Combien sommes-nous à consulter notre médecin et à acheter des médicaments sous l'influence de nos proches, des médias et de l'actualité ?

Selon Ipsos Public Affairs, Internet est devenu la seconde source de prises d'information médicale et d'auto diagnostic pour 67 % de la population interrogée. Quelle est l'influence d'Internet dans la part des remboursements de santé ? Quels sont les sujets de préoccupations des communautés de patients ? Comment et où positionner la prévention et le marketing digital en assurance Santé.

Cette session est un retour d'expérience de travaux menés en 2012 par OCTO pour le compte d'une société d'assurance, basés sur la collecte, l'analyse et la recherche de corrélations entre l'évolution des sujets d'intérêts médicaux et les dépenses de santé. Venez découvrir comment utiliser des techniques de traitements du langage naturel et de machine learning avec Hadoop permettent d'identifier des corrélations et des anticipations en matière de dépenses de santé sur le territoire français.

12h20 -13h : Open Data et Big Data : des cousins à marier ?

Par François LAINÉE, Président de CETADATA

Dans l'univers montant des Data, les segments Big et Open ont tous les deux la cote. On en parle, on en rêve, et on en fait un peu. Mais, si toutes les combinaisons de big/ small et open/closed sont en principe possibles, le caractère open porte en lui des forces particulières qu'il est intéressant d'analyser pour comprendre comment elles peuvent aussi affecter les Big Data.

Alors, quelles observations faire sur l'Open Data :

- *Tout d'abord son caractère encore très émergent.* En effet, selon toute mesure, il apparaît quasi anecdotique de par son étendue et son impact réel. Mais l'Open Data, c'est d'abord une révolution culturelle, et donc sa dynamique doit être appréciée à l'aune de la vitesse de ces phénomènes.
- *Malgré tout la diversité et les impacts des premières réalisations, témoignant de la puissance de la foule :* de nombreux exemples montrent en effet le pouvoir innovant et restructurant des analyses Data faites par la société civile
- *La puissance potentielle du mouvement, porteur de changement de la sphère publique :* Au fond, en permettant la mesure de politiques publiques, l'Open Data est d'abord un aiguillon douloureux dans la période de vaches maigres publiques. Mais il pourrait aussi devenir le socle d'un dialogue nouveau entre dirigeants et citoyens, un de ces ingrédients d'une future « big society ».

Cette force de la foule, support fondamental de l'Open Data, pourra jouer également dans le domaine des Big Data, soit en s'appuyant sur des Big Data qui pourraient bientôt s'ouvrir, soit en générant des Big Data « dirigées » par des projets de crowdsourcing à visée citoyenne ou sociétale.

A l'époque du mariage pour tous, les cousins Big et Open pourraient bientôt vouloir convoler eux aussi.

13h – 14h30 : Déjeuner

14h30 – 15h10 : S'appropriier les pouvoirs magiques du Big Data

Par Laurence HUBERT, PDG et CTO de HURENCE

Invariablement, les cas d'usage du Big Data régulièrement discutés et présentés tournent autour de la connaissance client, du marketing publicitaire et du suivi des campagnes, du sondage d'opinion et de la gestion d'image sur les réseaux

sociaux, de la détection de fraude. Tous ces usages sont réels et font sens mais leur listing aujourd'hui ressemble à un vaste "me too".

Le Big Data peut-il rendre innovant à lui seul? Tout semble indiquer que non.

Le Big Data est-t-il le meilleur allié de l'innovateur? Tout semble indiquer que oui. De Google à Facebook en passant par Twitter et Amazon, toutes les sociétés qui ont révolutionné notre espace numérique ces dernières années en ont fait leur meilleur allié.

Le thème de ce talk sera de démontrer que le Big Data a bien des pouvoirs et en aura encore d'avantage si on a l'ambition de se les approprier à travers lui.

15h10 – 15h50 : Le Big Data avance à grands pas : un exemple dans les Smart Grids

Par Dr. Michael DEFOIN-PLATEL, Senior Data Scientist chez GRIDPOCKET SAS.

Les Smart Grids, réseaux intelligents de production, transport, distribution et consommation d'énergie, sont déjà une réalité pour l'électricité aux US alors que des projets commencent à se concrétiser en Europe et en Asie. Les efforts se sont pour l'instant portés principalement sur le déploiement d'équipements intelligents de mesure et de contrôle en différents points de la grille ainsi que sur l'installation de réseaux de communication pour le pilotage de ces équipements et la collecte des données. Parmi les équipements déployés, les compteurs intelligents, qui seront à terme présents dans tous les foyers, sont capables d'effectuer des relevés de consommation à des fréquences bien supérieures à ce qu'il était possible de faire avec les installations classiques (de une à plusieurs dizaines de mesures par heure).

Les prochaines étapes pour la mise en place des Smart Grids requièrent le développement de solutions capables de stocker, d'intégrer mais également d'analyser et de fouiller les données issues de ces grilles pour mieux gérer leurs différents composants et fournir des services qui facilitent la compréhension et la maîtrise de la consommation électrique. La société GridPocket, pour anticiper les volumes colossaux de ces données et la complexité des tâches à réaliser, s'est positionnée sur une approche Big Data basée sur la technologie Hadoop pour

concevoir une plateforme de services qui puissent couvrir les besoins des utilisateurs finaux qu'ils soient résidentiels, industriels ou fournisseurs d'énergie. Pour accroître encore son avance dans ce secteur, GridPocket est membre du projet Européen BigFoot (<http://bigfootproject.eu/>) qui ambitionne d'améliorer l'écosystème Hadoop et de faciliter son utilisation industrielle, que ce soit au niveau du stockage des données, mais également au niveau du développement, du déploiement, et de l'exécution d'applications distribuées.

Dans cette présentation, nous détaillerons dans un premier temps les différents défis posés par les Smart Grids et les réponses Big Data proposées par GridPocket. Dans un deuxième temps nous évoquerons les limitations des technologies actuelles et leurs conséquences pour les activités de GridPocket. Enfin, nous discuterons des objectifs du projet BigFoot et des approches suivies pour faire avancer le traitement des données massives.

15h50 – 16h10 : Pause café

16h10 – 16h50 : Enjeux et Perspectives du Big Data : application au géomarketing dynamique :

Par Mathieu GRAS, Responsable Services Géolocalisés chez SFR

Au cours de cette présentation seront abordés les thèmes suivants :

- les enjeux du Big Data pour SFR ou comment répondre à la fois à un besoin interne et externe.
- une application du Big Data : le geomarketing dynamique

17h – 17h30 : Big Data, Big Brother ?

Par Didier TRANCHIER, PDG d'ADELIT

L'avènement du Big Data va créer une révolution complète de notre économie et de notre société, allant des technologies aux usages, des Business Models au respect des individus. C'est à la fois autant d'opportunités et de menaces pour les acteurs de cette économie, que surtout pour l'utilisateur final en termes de : SECURITE, MORALITE, et de PROTECTION des BIENS et des LIBERTES.

