

De la Data aux Big Data : enjeux pour le Marketing client - Illustration à EDF

ENTRETIEN AVEC ANNE GAYET ET JEAN-MICHEL GAUTIER



Jean-Michel GAUTIER

Professeur des départements Economie & Marketing à HEC, en charge de la Connaissance Client à EDF Commerce (commercialisation et gestion des contrats de fourniture des différents marchés : particuliers, professionnels, entreprises et collectivités locales).

Anne GAYET

Directrice associée chez AID (Add Intelligence to Data). AID accompagne les équipes de la Direction Numérique d'EDF-Commerce. Par ailleurs AID propose des services d'hébergement de CRM marketing, de datamining / data science, d'analyse des parcours clients et de conseil pour le marketing opérationnel.

Quand avez-vous débuté avec le "Big Data"?

Jean-Michel Gautier

Dès 2003, date de sa création, EDF Commerce a été confronté, à travers l'ouverture progressive des marchés, à l'arrivée de concurrents, et à la nécessité de passer d'une approche usager à une approche client. C'est dans ce contexte que se sont construites progressivement la connaissance client et les plateformes de stockage et d'analyse des données clients.

Après une première étape de développement d'entrepôts client et la mise en place d'une activité d'analyses au service du marketing et des opérations, s'est posée la question de l'adoption de nouvelles technologies capables de supporter de l'interaction en temps réel et d'exploiter les nouvelles sources de données que constituent le web, les applications mobiles, les objets connectés et en particulier les compteurs intelligents.

Nous avons décidé dès 2012 de migrer l'ensemble des infrastructures des entrepôts de données vers une architecture Hadoop, avec un double objectif fonctionnel et économique. Nous avons en parallèle migré les activités d'analyse vers l'open source R, en remplacement de SAS.

Anne Gayet

Chez AID nous avons pendant ce temps démarré une transformation complète de l'approche des "parcours clients".

Nous avons décidé en 2013 de développer une nouvelle façon d'analyser et de gérer les parcours clients, en développant une plateforme "Big Data" pour la centralisation de toutes les interactions clients quel que soit le canal, pour leur structuration, et pour l'analyse interactive des parcours. Entièrement élaborée à partir d'outils open source, notre solution a été initiée autour de deux ensembles d'outils : Hadoop et Hbase (une base de données NoSQL orientée colonnes) pour le stockage et l'analyse des données d'une part, Kafka pour la gestion de la file d'attente et Storm pour l'intégration et la transformation des données au fil de l'eau d'autre part. Une interface de visualisation totalement originale a été développée.

Qu'en est-il aujourd'hui? Les outils utilisés ont-ils évolué?

Jean-Michel Gautier

Cette rupture complète des technologies de stockage et d'exposition de données, de DB2 et Oracle vers Hadoop, Hbase et de SAS vers R, est un succès total avec un triple bénéfice :

- Réduction à périmètre égal des coûts de fonctionnement des entrepôts
- Capacité de stockage étendue à très bas coûts
- Passage progressif vers le traitement des interactions client en temps réel (en particulier

sur le web)

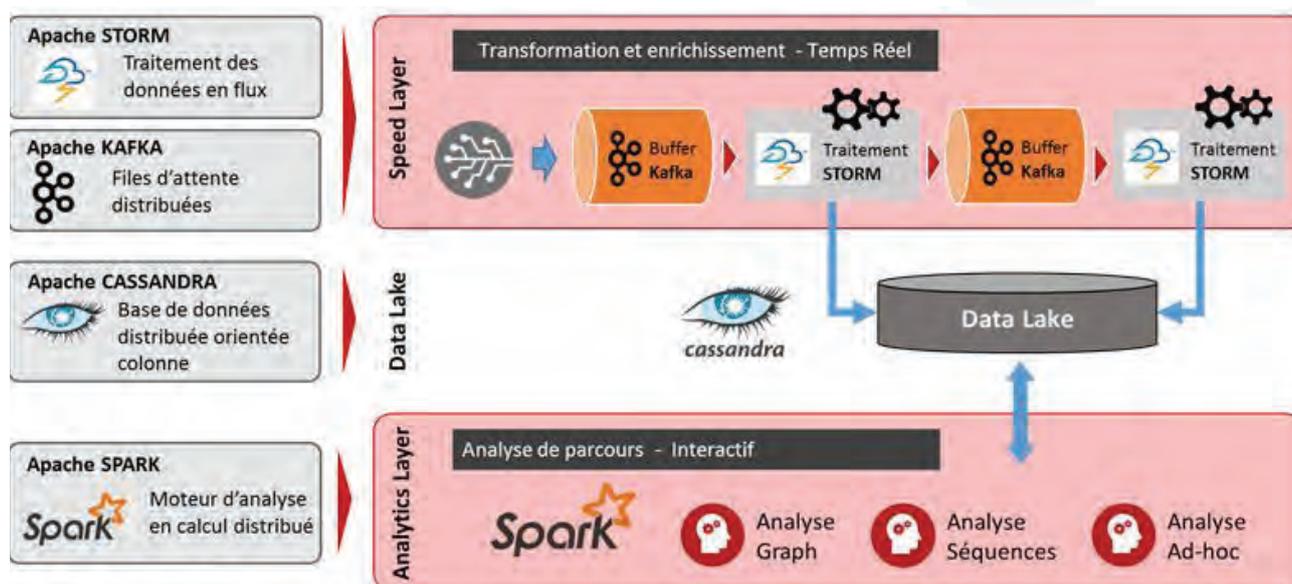
L'utilisation de Hbase pour l'exposition de données en temps réel et de Spark associé à R, est la suite logique pour travailler efficacement dans un environnement distribué.

Par ailleurs de nombreux tests d'outils autour d'Hadoop sont en cours.

Anne Gayet

Nous avons initialement pensé notre outil avec Hadoop et Hbase pour la partie analytique, mais nous avons rapidement migré vers Spark et Cassandra :

- Spark pour sa plus grande vélocité d'analyse réalisée en mémoire, et la richesse de sa bibliothèque d'algorithmes en calculs distribués (MLlib : Machine Learning library).
- Cassandra pour sa facilité d'utilisation, notamment grâce à son langage de requête CQL, qui ressemble au SQL.



Copyright **Aid** 2016

Qu'est-ce qui a fait les premiers succès auprès des métiers et en particulier des marketeurs?

Jean-Michel Gautier

La réduction des coûts de stockage sur Hadoop et la facilité à acquérir de nouveaux flux de données ont complètement changé la vision de la gouvernance de la donnée.

Avant, il fallait démontrer la valeur d'usage pour justifier les dépenses de collecte et stockage. Aujourd'hui on peut capter et collecter tout ce qui passe sans le traiter. La modélisation objet, le traitement et l'analyse se font au fil des besoins selon les cas d'usage. L'investissement data se fait donc au fur et à mesure en fonction du business model des cas d'usages.

Chez EDF, je relèverai quatre facteurs clés de réussite :

- La rapidité à délivrer une solution isopérimètre (15 mois avec livraison au fil de l'eau par périmètre fonctionnel).
- Les premiers cas d'usage d'exposition de données en temps réel pour de nouveaux services au client : par exemple la possibilité pour les entreprises multi-sites d'accéder à leurs consommations et factures via le web et de regrouper à volonté leurs données selon des logiques d'agrégation diverses de leurs sites.
- La continuité, l'extension des activités décisionnelles classiques avec l'analyse de données non structurées (réclamations, commentaires conseillers, aides en ligne, mails, tweets), l'analyse fine des consommations (issues des compteurs Linky) à des fins d'accompagnement à la maîtrise énergétique.

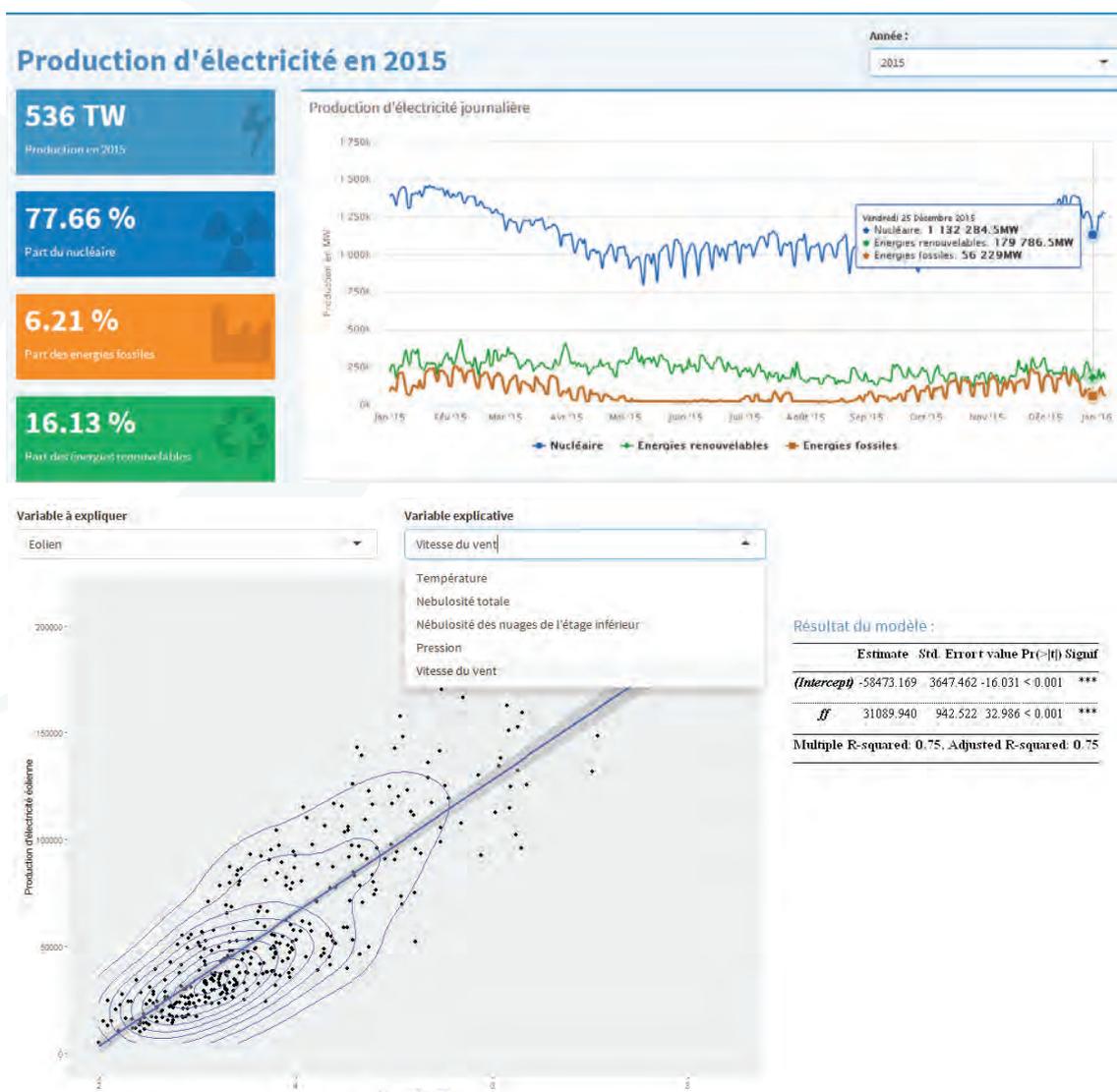
- La capacité à montrer en interne de nouvelles applications illustrant l'efficacité et la pertinence d'un traitement de bout en bout en temps réel. Par exemple : la collecte, le filtrage et la catégorisation en temps réel des tweets qui concernent EDF selon des axes prédéfinis visibles pour les managers concernés.

Anne Gayet

La dimension datavisualisation est extrêmement importante, tant chez EDF que dans notre plateforme datakili®. Cette dimension se concrétise de plusieurs façons :

- La restitution et l'analyse interactive avec des outils comme Tableau Software, mais aussi en open source avec R Shiny, ou D3.js.
- La synthèse des éléments de connaissance clients apportés par les analyses, sous forme d'infographies, de présentations très « design » et visuelles, ou encore par des mini-vidéos ou animations. Ces modes là n'ont rien de « Big Data » ...
- Et bien sûr par le développement d'interfaces orientées « dataviz » quand il s'agit d'outils ou d'applications.

A titre d'illustration ci-dessous un exemple d'application Shiny très simple permettant d'analyser la production d'électricité au cours du temps, et en fonction de paramètres météorologiques (sources de données ouvertes).



Nb : De nombreux exemples d'applications sont consultables dans ce catalogue en ligne <http://www.showmeshiny.com>

Quelles ont été les difficultés à surmonter ?

Jean-Michel Gautier

Dans l'exploitation opérationnelle des Big Data, le premier besoin est la capacité à exposer et traiter de l'information en temps réel, ce qui est relativement simple avec l'écosystème Hadoop. Même si, comme beaucoup d'entreprises, EDF est encore contraint par ses systèmes d'information traditionnels dont l'interface de publication est quotidienne, en nuit applicative. Une migration complète des politiques de publication de données par les SI traditionnels est nécessaire pour arriver au temps réel.

Cette contrainte aboutit aujourd'hui à une gestion du temps réel canal par canal, le plus facile étant les canaux web et mobile. La difficulté et l'enjeu à venir est bien la gestion client omnicanale en temps réel.

Anne Gayet

Une des plus grandes difficultés que nous rencontrons toujours chez nos clients est la lutte contre les « silos » d'informations. Par exemple il est encore fréquent que les comportements de navigation sur les sites web de l'entreprise ne soient pas ou peu accessibles par les analystes. Une solution consiste à poser un tag spécifique sur les sites web de l'entreprise pour alimenter les bases de données. Mais la maîtrise de la technique permettant de tracer la navigation sur un site web ou l'utilisation d'une application mobile est une compétence très particulière que n'ont pas les data managers et dataminers. La solution souvent pratiquée consiste à utiliser un traceur ou un gestionnaire de tag déjà implémenté, à condition que ce système donne accès aux données de navigation détaillées et pas uniquement à des comptages (par exemple Google Analytics Premium, Adobe Analytics, Tag Commander, ...).

C'est quoi le marketing chez EDF Commerce ?

Jean-Michel Gautier

C'est avant tout une dimension offre d'électricité et de services (conseil tarifaire, rénovation) couplée à un programme relationnel, une fidélisation par la relation client et une attention portée à tous les événements clés de la relation client : déménagement, changement de puissance, difficultés de paiement, travaux. Ces prestations sont axées sur le confort dans l'habitat pour les particuliers et sur le service commercial pour les entreprises, mais aussi pour tous sur l'accompagnement de la maîtrise de l'énergie.

Pouvez-vous nous citer des cas d'usage en cours de développement ou à venir ?

Jean-Michel Gautier

Nous avons déjà évoqué l'analyse de tweets. Notre « Tweet Tracker » n'est pas uniquement destiné à prouver la possibilité de capter - analyser - restituer en temps réel. Il intègre des fonctions d'analyses linguistiques poussées non disponibles dans les plateformes de veille des réseaux sociaux. C'est le précurseur d'outils de relation client capables de comprendre la parole client.

De nouveaux services tirant parti d'objets connectés sont expérimentés : à partir de la consommation client et des températures (celle du logement, la température extérieure) on propose un service de régulation du chauffage.

Nous disposons de différentes zones d'expérimentation poussées, dont Smart Electric Lyon et Nice Grid, où le foisonnement technologique est de mise.

Le fait d'utiliser Hadoop permet d'ouvrir plus aisément les données à l'extérieur du SI. Les données sont ainsi présentées au client dans son espace client. Elles sont également poussées vers le conseiller en ligne qui bénéficie ainsi d'une « vision 360° » du client, pour une connaissance

parfaite de son interlocuteur et une relation commerciale optimisée. Les pistes d'investigation actuelles s'appuient parfois sur des startups, par exemple en ce qui concerne la mise en place de nouveaux portails de visualisation pour les métiers ou partenaires.

Anne Gayet

L'analyse des parcours clients, ou le domaine appelé « expérience client », s'appuie habituellement sur une optimisation des processus de gestion de la relation client ou sur une mesure par des enquêtes. Une solution comme datakili® permettant de stocker, quantifier et analyser tous les parcours clients permet de les concrétiser, les analyser et faire de la prédiction. On peut ainsi par exemple identifier les événements précurseurs de la résiliation, mieux comprendre les parcours cross-canaux ou encore mettre en évidence les séquences d'interactions client - entreprise menant à un appel au call center.

Pour conclure qu'apporte ce nouvel écosystème Big Data?

Anne Gayet

En ce qui concerne les capacités analytiques pour le marketing, les solutions traditionnelles étaient adaptées à ce fonctionnement : compréhension du besoin marketing, compréhension des données, collecte et préparation des données, modélisation prédictive, évaluation, déploiement, surveillance et mises à jour. D'abord l'écosystème « Big Data » permet de satisfaire ces besoins classiques de façon plus performante en rapidité, en qualité statistique et en pertinence vis-à-vis du client. C'est à l'heure actuelle toujours le principal apport. Cette meilleure performance est obtenue grâce :

- à la baisse des temps de calcul,
- des mises à jour fréquentes des modèles,
- la prise en compte de données plus variées et détaillées pour une meilleure capacité explicative, cela inclut des données externes comme la météo,
- l'utilisation possible en temps réel pour être au plus près du « temps du client »,
- à l'application d'algorithmes très gourmands en calculs (parfois).

Par ailleurs, ce nouvel écosystème permet de diffuser l'utilisation des données au service des métiers, via la mise en place de « hubs de données » associés à des outils permettant aux utilisateurs d'explorer et visualiser les données en leur masquant la complexité sous-jacente.

Jean-Michel Gautier

Nous avons évoqué des développements internes à EDF, mais les startups de l'écosystème sont également des sources potentielles d'accélération.

Nous organisons donc une veille active et procédons à des tests de nouvelles technologies ou de nouveaux services, via des POCs (proof of concept) réalisés en interne ou en externe, i.e. des expérimentations limitées permettant de faire la preuve du concept avant le déploiement. Le marché de l'énergie est en pleine mutation : énergie verte, nouveaux usages de l'électricité (voitures électriques), production locale (client fournisseur), équilibre de réseaux, etc. Et tout ceci produit un foisonnement de données qu'il faut savoir collecter et traiter pour des services à inventer. Seules les technologies du Big Data nous rendront capables d'aborder ces enjeux.