

Mai 2017

Data Analytics : de la visualisation au prédictif



En bref

La Matinée Stratégique « Data Analytics : De la visualisation au prédictif » a été organisée par CIO le 25 avril 2017. Organisée en partenariat avec Axway, Informatica et HPE Vertica, avec le soutien de Forrester, elle a permis le témoignage d'entreprises comme Mappy, Kaufman et Broad, le CEA, Chronopost, Ouibus, Criteo et Air France. Forrester a également apporté son éclairage.

Les interventions des uns et des autres sont ici synthétisées.

Sommaire

Stratégie

Data Analytics : savoir visualiser et prédire

Stratégie

Henry Peyret (Cabinet Forrester) souligne l'importance des systems of insight et de la datagouvernance

Projets

Comment visualiser ses données avec efficience ?

Projets

Chez Chronopost, la technologie traite de la data pour le business

Projets

Le prédictif au service du pilotage business

Data Analytics : savoir visualiser et prédire



Managers IT ou spécialistes du décisionnel ont assisté à la Matinée Stratégique CIO « Data Analytics : De la visualisation au prédictif » le 25 avril 2017.

La Matinée Stratégique « Data Analytics : De la visualisation au prédictif » a été organisée par CIO le 25 avril 2017. Organisée en partenariat avec Axway, Informatica et HPE Vertica, avec le soutien de Forrester, elle a permis le témoignage d'entreprises comme Mappy, Kaufman et Broad, le CEA, Chronopost, Ouibus, Criteo et Air France.

En partenariat avec Axway, Informatica et HPE Vertica, CIO a organisé une Matinée Stratégique « Data Analytics : De la visualisation au prédictif » le 25 avril 2017 au Centre d'Affaires Paris Trocadero. Mappy, Kaufman et Broad, le CEA, Chronopost, Ouibus, Criteo et Air France y ont témoigné de leurs démarches.

En ouverture de la conférence, la rédaction de CIO a dévoilé en avant-première les principaux résultats de l'étude [Quelles connaissances disponibles au-delà du Big Data ?](#). Celle-ci a montré que les meilleures pratiques sont loin d'être universellement adoptées dans les entreprises françaises.

Importance de l'infrastructure pour une bonne qualité de service



François Guerin, PreSales Consultant Software Big Data chez HPE Vertica, a plaidé : « Optimisez l'analyse de vos données en un clin d'oeil et en toute agilité »

Mais, pour profiter pleinement des nouvelles approches et technologies destinées à mieux exploiter les données, il ne faut pas négliger le rôle et l'importance des infrastructures sous-jacentes. Celles-ci conditionnent les fonctionnalités disponibles et la qualité de service, comme l'a rappelé François Guerin, PreSales Consultant Software Big Data chez HPE Vertica. Une solution universelle couvrant tous les besoins reste, par contre, à ce jour un fantasme. François Guerin a ainsi relevé : « souvent, on oppose des technologies comme Hadoop au SQL alors qu'il faut les intégrer en écosystème. Le véritable intrus est bien la solution universelle. Vertica est un socle assurant un support complet de SQL, avec une architecture massivement parallèle, un stockage en colonne et une compression native mais il faut utiliser la bonne technologie pour le bon usage en maintenant les coûts dans des limites raisonnables. »



Henry Peyret, Analyste Principal de Forrester Research, a détaillé les outils et méthodes pour visualiser et prédire à partir des données.

La gouvernance des données stockées dans les infrastructures doit elle aussi évoluer. Au delà de la seule préparation des données -un sujet en lui-même-, cette gouvernance doit permettre de passer à une nouvelle phase du marketing pour s'adapter aux nouvelles exigences des consommateurs. « Une technologie de type Datawarehouse n'est plus de mise » a affirmé Henry Peyret, Analyste Principal de Forrester Research. Malgré tout, les entreprises peuvent être soumises à une grande tentation de traiter n'importe comment les données. Pour Henry Peyret, il ne fait donc aucun doute que « nous aurons des marketeurs fous comme nous avons eu des traders fous. Ce sont d'ailleurs les mêmes technologies qui sont à l'oeuvre pour traiter les données. » Il faut donc toujours se rappeler les finalités et travailler en mode de boucle de rétro-action

pour garder la technologie au service des utilisateurs, avec le bon niveau de qualité.

Etre au service du Parcours Client



Edouard Guérin, Senior Sales Consultant chez Informatica a expliqué comment « Optimiser le parcours client en temps réel grâce au Big Data ».

Spécialisé en collecte de données, gestion de leur qualité et master data management (MDM), Informatica se place directement dans ce fil conducteur. Pour servir les clients et les autres utilisateurs, l'approche doit être à la fois Big Data et temps réel, avec un traitement de données au sein de data lakes. Cette approche seule permet d'optimiser un parcours client global constitué de multiples expériences qui sont elles-mêmes des mini-parcours-clients. « Le Parcours Client est une somme de parcours clients comme l'Histoire est une somme d'histoires » a plaidé Edouard Guérin, Senior Sales Consultant chez Informatica.



La première table ronde, « Visualiser les données avec efficacité », a réuni, de gauche à droite, Nicolas Korchia (Responsable Business Intelligence chez Mappy), Philippe Minier (DSI de Kaufman et Broad) et Damien Poulain (architecte applicatif au CEA).

Première série de témoignages d'entreprises, la première table ronde était consacrée au thème « Visualiser les données avec efficacité ». Elle a réuni Nicolas Korchia (Responsable Business Intelligence chez Mappy), Philippe Minier (DSI de Kaufman et Broad) et Damien Poulain (architecte applicatif au CEA). Ceux-ci ont détaillé leurs approches techniques et business.

Accompagner la transformation des habitudes des consommateurs



« Améliorer l'expérience client grâce aux analytics temps réel » a été le thème détaillé par Said Debbagh, Product Marketing Director & Expert - Advanced Analytics chez Axway.

Parmi les bouleversements que doivent accompagner les DSI, il y a bien sûr la transformation digitale qui est cause et conséquence d'évolutions importantes dans les habitudes des consommateurs. Celle-ci est liée à une augmentation des données générées mais aussi à une accélération des échanges et des traitements. L'entreprise doit comprendre les comportements de ses clients et anticiper leurs évolutions. « Un problème devient un vrai problème quand il est constaté par l'entreprise après le client » a ainsi soutenu Said Debbagh, Product Marketing Director & Expert - Advanced Analytics chez Axway. S'appuyant sur des cas de plusieurs entreprises, Said Debbagh a voulu insister sur un conseil fondamental : « l'approche doit être itérative ». Il est en effet impossible de tout savoir, surtout ce qui change en continu.

Enrichir le service par les données



Le Grand témoin de la Matinée, Hugues de Maussion, DSI de Chronopost, a exposé les usages du Big Data et de l'Internet des Objets au service du business.

Intervenant au fil de la matinée, notamment pour conclure les tables rondes, le Grand Témoin était Hugues de Maussion, DSI de Chronopost. Il a, au cours d'une interview d'une vingtaine de minutes, présenté les réalisations de son entreprise qui a su utiliser avec pertinence et innovation les données, y compris l'IoT, pour améliorer le service client et lancer de nouvelles offres.



La deuxième table ronde, sur « Le prédictif pour piloter l'activité » a réuni, de gauche à droite, Béatrice Godineau (Directrice Marque et expérience client de Ouibus), Bastien Albizzati (Product Director Engine & Marketplace chez Criteo), Isabel Gomez Garcia De Soria (Directrice de la Recherche Opérationnelle chez Air France KLM).

Enfin, la matinée a été conclue par une seconde table ronde consacrée au sujet « Le prédictif pour piloter l'activité ». Elle a réuni Béatrice Godineau (Directrice Marque et expérience client de Ouibus), Bastien Albizzati (Product Director Engine & Marketplace chez Criteo), Isabel Gomez Garcia De Soria (Directrice de la Recherche Opérationnelle chez Air France KLM).



Comme toujours la matinée s'est conclue par un cocktail permettant non seulement de discuter entre managers IT mais aussi de visiter les stands des partenaires.

En savoir plus

- [Contenus associés à la conférence Data Analytics du 25 avril 2017](#) (Progressivement enrichis).
- Télécharger l'étude [Quelles connaissances disponibles au-delà du Big Data ?](#)

Bertrand Lemaire
Rédacteur en chef de CIO

Henry Peyret (Cabinet Forrester) souligne l'importance des systems of insight et de la datagouvernance



Henry Peyret, Principal Analyst chez Forrester, base ses recherches sur la datagouvernance.

En s'appuyant sur les témoignages de DSI (Mappy, Kaufman and Broad, le CEA, Critéo, Air France KLM, Ouibus), CIO a organisé le 25 avril 2017 une Matinée Stratégique «Data Analytics, de la visualisation au prédictif ». La conférence a été réalisée en partenariat avec Axway, Informatica et Vertica. Henry Peyret (Cabinet Forrester) y a apporté son expertise en ouverture de conférence.

« Je suis analyste depuis 17 ans, je m'occupe de datagouvernance, une datagouvernance adaptée aux nouveaux usages, lance Henry Peyret, Analyst Principal au cabinet Forrester, qui ouvrait la Matinée Stratégique CIO «Data Analytics, de la visualisation au prédictif », le 25 avril dernier. En fait, ce que l'on est en train de découvrir ce sont de nouveaux utilisateurs et de nouveaux usages. Les deux autres axes de recherche sont les problématiques d'intégration, notamment celle des données, et la préparation de la data, le data wrangling. Troisième sujet que vous allez voir apparaître dans les 10 ou 15 prochaines années, c'est en fait une nouvelle classe de consommateurs, des consommateurs qui vont consommer en fonction de leurs valeurs. Ils vont changer, utiliser de nouveaux outils, ce que l'on appelle les réseaux sociaux 2.0, de nouveaux réseaux sociaux pour consommer différemment. Cela va générer une nouvelle approche en phase avec le marketing 3.0. »

Henry Peyret essaie de comprendre et de faire comprendre quelle est la place de ces nouvelles technologies pour ces nouveaux utilisateurs et pour ces nouveaux usages. « On a utilisé le datawarehouse pendant des années, et on s'aperçoit que ce ne sont pas ces technologies, de type datawarehouse, qu'il faut pour les nouveaux usages. Je vais

vous montrer avec les systèmes of insight des architectures un peu différentes. Donc qu'est-ce qui fait que ces systèmes sont différents des systèmes datawarehouse ? »

L'intérêt des systems of insight

D'abord, il faut reconnaître qu'on a fait depuis quelques années une séparation entre les systèmes of records et les systèmes of engagements qui touchent les consommateurs et les employés, remarque Henry Peyret. « On voit apparaître aussi un autre système, les system of automation, qui vont connecter avec tous les objets et le monde physique, par exemple le smart grid et tout ce qui va accompagner le smart grid. On voit apparaître un system of insight qui, autour du datalake, où on met toute la donnée, va permettre l'exploitation de la donnée et de ses différentes formes. Les premières utilisations de ces systems of insight, c'est bien entendu de permettre de nouvelles actions, pour les system of engagement. On va trouver d'autres utilisations à l'avenir, les system of insight liés aux objets connectés et même l'interconnexion entre les objets connectés et les systems of engagement. On va voir également de plus en plus d'interconnexion avec ceux qui fabriquent quelque chose, par exemple des avions, c'est les system of design. »

Qu'est-ce que c'est que ces systems of insight et qu'est ce qui change avec eux ? Dans les systems of engagement, par rapport au datawarehouse, il y a deux changements fondamentaux. Selon Henry Peyret, on va d'abord chercher à détecter des insights marketing et commerciaux pour des actions quasi immédiates, en tout cas très rapides, dans une heure, pour faire une campagne marketing, la rectifier et la vérifier rapidement, c'est un changement fondamental dans le monde du marketing. « Avant, on obtenait les résultats un mois, trois mois après. A l'heure d'Internet c'est un quart d'heure après, donc vous pouvez prendre des décisions, cela très vite. Ce system of engagement doit nous dire si nous sommes dans le bon insight. On génère de nouvelles données notamment pour voir si l'action et l'insight sont les bons. »

La boucle de feedback

Quand on va un peu plus dans le détail, on s'aperçoit que, par rapport aux actions faites dans le monde de la BI ou du datawarehouse on peut aboutir à des actions supplémentaires. Il faut d'abord identifier les métriques et la manière de ramener des données supplémentaires afin de vérifier que ça a du succès. « Ensuite, on va collecter plus de données, développer des insights qui vont générer des actions et tester et implémenter ces insights, ce sera peut-être pas les bons, peut-être pas les bonnes métriques, mais c'est cette fameuse boucle, en utilisant l'intelligence artificielle, qui va prévaloir. Ce que j'essaye de faire comprendre avec mes recherches sur la datagouvernance, c'est qu'on a intérêt à faire valoir cette fameuse boucle de feedback. »

« Attention à ne pas trop embêter le consommateur qui dispose d'outils suffisants (on pense aux adblockers pour contrer les publicités) lui permettant d'échapper aux annonceurs. C'est une manière pour le consommateur de se protéger. Si on veut qu'il vous fournisse suffisamment d'informations pour avoir cette boucle de feedback, il ne faut pas embêter le consommateur. Je vous engage dans les systèmes d'insight que vous allez mettre en place à changer votre datagouvernance. Les consommateurs vont trouver des outils et ça fera très mal. Si vous ne comprenez pas et ne mettez pas des garde-fous à vos marketeurs, vous aurez des marketeurs fous comme il y eut des traders fous. »

Des états de maturité très variables

Un system of insights, c'est un ensemble de technologies, pas une seule technologie. Il y a des états de maturité très variables, avec des technologies en pleine évolution, de même pour les technologies en mode croissance qui vont contribuer à construire les systèmes of insight. On est plus du tout dans la construction d'un datawarehouse, on est dans un mode itératif.

« Quelques recommandations pour finir. Un, le principe même des systems of insights c'est, non seulement de fournir des actions vers lesquelles on doit aller, mais c'est aussi de fournir la final loop, c'est-à-dire le moyen de vérifier la portée de cette action. Deux, c'est de construire le system of insight en ayant conscience qu'il y a énormément de technologies, non seulement aujourd'hui mais à l'avenir, qu'il y a aussi besoin parfois de dupliquer et rationaliser ces technologies, mais on est très loin d'être dans les tables de rationalisation aujourd'hui, on commence à rationaliser à partir du moment où ce seront des technologies relativement matures, on en est très loin aujourd'hui. »

« Troisième élément qui me paraît fondamental, penser à importer de la datagouvernance. Gouvernez véritablement vos systèmes of insight, il y a beaucoup d'écueils à laisser les data scientists et les data analysts faire leurs choix dans leur coin. Il y a énormément de choses mais aussi de bêtises à faire. Et des bêtises, on va en voir de nombreuses dans les prochaines années, si on en cadre pas suffisamment ces data analysts et ces data scientists. »

En savoir plus

- [Contenus associés à la conférence Data Analytics du 25 avril 2017](#) (Progressivement enrichis).
- Télécharger l'étude [Quelles connaissances disponibles au-delà du Big Data ?](#)

Didier Barathon
Journaliste

Projets

Comment visualiser ses données avec efficacité ?



Lors de la table ronde sur la datavisualisation, de gauche à droite : Nicolas Korchia, Mappy, Philippe Minier, Kaufman and Broad, Damien Poulain, CEA.

CIO a organisé le 25 avril 2017 une Matinée Stratégique «Data Analytics, de la visualisation au prédictif ». Philippe Minier, DSI de Kaufman and Broad, Damien Poulain, architecte applicatif au CEA et Nicolas Korchia, responsable business intelligence chez Mappy, ont ainsi participé à une table ronde sur la visualisation de la donnée.

Comment visualiser ses données avec efficacité ? C'était le thème de la première table ronde lors de la Matinée Stratégique sur « Data Analytics » que CIO organisait le 25 avril dernier. Table ronde donnant la parole à trois managers IT : Philippe Minier, DSI de Kaufman and Broad, Damien Poulain, architecte applicatif au CEA et Nicolas Korchia, responsable business intelligence chez Mappy.

Philippe Minier, devait remettre le data analytics dans son contexte : respecter les obligations de qualité de service, le besoin de gérer les fournisseurs, de facturer. *« La qualité est un point fort chez nous, c'est ce qui a fait la marque. Rechercher le zéro défaut est très compliqué, mais on a mis en place des outils de dernière génération, par exemple pour assurer la livraison d'un bien au client. L'un des points importants pour nous en termes de qualité, c'est de minimiser le nombre d'incidents lors des livraisons et ensuite aller très vite pour régler les petits défauts qui peuvent rester. »*

« On a mis en place des outils agiles avec de la tablette tactile, de la récupération de données. On récupère de plus en plus de données du côté des fournisseurs, des prestataires et derrière on trace tout ce qui se passe en termes de levées de réserve, de réactivité des entreprises. Derrière, il y a beaucoup d'automatisation avec tous ces outils, on récupère de plus en plus de données, et on les traite, on les remonte à la direction générale avec des indicateurs. On fait de l'analyse très « high level » mais on est capable ensuite de descendre vers du plus fin grâce aux technologies Qlick. »

Assurer tout le cycle de vie d'une opération

Le spécialiste de l'immobilier est capable de faire de l'agrégation mais aussi de faire descendre les données Qlick, Tableau, Vertica, Alteryx, pour ses différents besoins, sur la partie livraison. La qualité c'est plutôt Qlick. Vertica étant plus pour le suivi financier très fin par opération, avec des fiches d'opérations, du Tableau dessus pour faire de la dataviz. Tout le cycle de vie d'une opération immobilière est assuré par ces outils de data analytics, en partant de la recherche foncière, très importante. Un des problèmes aujourd'hui, c'est le foncier, la réservation et son cycle de vie, et le montage de l'opération. *« On utilise les nouvelles technologies pour améliorer ce suivi complet. On a pas trouvé une solution qui fasse tout, on est obligé d'agglomérer plusieurs solutions. Derrière, on capitalise les données, les observations faites sur un chantier, tous les indicateurs sont réutilisables par exemple pour identifier les meilleurs fournisseurs et leur évolution. Ce qui sert aux appels d'offres suivants. »*

Pour Damien Poulain, du CEA, *« on a utilisé au fil du temps une grande variété d'outils décisionnels, des données classiques, on essaie de limiter les outils par technologie, un en base de données, un seul en BI, Business Object en BI objet en BI agile on vient de valider Digidash. On a une hétérogénéité d'activités, ce n'est pas une raison suffisante pour être hétérogène en outils. On essaie d'optimiser en ayant le meilleur outil pour chaque technologie».*

« Je prône une BI agile qui aille jusqu'au bout »

Quelle est la limite entre Business Objet et Digidash ? Dans l'agilité il y a tout, l'intuitivité, la capacité à fabriquer de la BI sans formation, sans prestataire de service. *« Je prône une BI agile qui aille jusqu'au bout. L'utilisateur non informaticien peut s'en servir avec un collègue qui montre comment ça marche, ça reste simple, pas de coût supplémentaire. La BI traditionnelle avec du datawarehouse, a fait ses preuves, elle n'est pas obsolète pour accéder aux données. Le standard BO existe. Le but c'est qu'on intervienne le moins possible, je ne sais même pas quels sont les tableaux de bord, on a mis la plateforme, on assure la disponibilité, la performance, l'infrastructure. Par contre, l'usage nous échappe, mais c'est ce qu'on souhaite, que ce soit le moins bloquant dans le processus. Je pense qu'Excel est l'outil de la BI historique, avec Digidash on n'est pas loin de l'usage qu'on a avec Excel. C'est quand même la première fois que je vois un outil capable de concurrencer Excel pour le tableau de bord (pas pour le côté tableur évidemment). »*

Nicolas Korchia, de Mappy, a, lui, utilisé successivement beaucoup d'outils décisionnels et de traitements de la donnée, il était l'un des premiers sur Hadoop. Son gros sujet, c'est l'analyse des logs, pour vérifier que les itinéraires proposés sont bons. L'équipe BI travaille pour comprendre ce que font les utilisateurs et comprendre comment sont utilisés les produits Mappy et cela en se basant les logs. 13 millions de visiteurs uniques chaque mois, 500 millions de requêtes par mois : mais il faut sortir l'information essentielle. Sur certains logs, on ne cherchera pas d'information supplémentaire. Mais, sur d'autres, au contraire, il faudra descendre de manière très fine. Mappy essaye donc de structurer au maximum la donnée pour que l'utilisateur final soit à l'aise, même si on lui met à disposition un maximum de données.

L'utilisateur a aussi un niveau d'exigence important

« On est autour de 10 To de données utiles chez Mappy, de la donnée qui contient l'historique de tous les itinéraires ou bien l'audience de chacun des centres d'intérêt. 2 ou 3 To supplémentaires arrivent chaque semaine. Nous utilisons Tableau pour la visualisation. On a croisé une start up Indexima, avec des temps de réponse

intéressants, à la milliseconde. L'idée c'est d'être transparent, que l'utilisateur ne se rende pas compte qu'il est en lien avec des données hallucinantes. L'utilisateur a aussi un niveau d'exigence important. Mes utilisateurs sont très durs, et on a 5 milliards de lignes. »

La qualité de la donnée reste une grande préoccupation. « *On a fait des progrès, on voyait des anomalies dans les graphiques, on amélioré la qualité, souligne Philippe Minier, aujourd'hui on est plus dans le redressement. On utilise Alteryx pour gérer cette partie au moment de l'intégration pour nettoyer la donnée* ».

Au CEA, remarque Damien Poulain, on est moins en avance mais il n'y a pas non plus un grand besoin. Conclusion générale : On va avoir de plus en plus de données, c'est exponentiel, c'est cela le grand sujet.

En savoir plus

- [Contenus associés à la conférence Data Analytics du 25 avril 2017](#) (Progressivement enrichis).
- Télécharger l'étude [Quelles connaissances disponibles au-delà du Big Data ?](#)

Didier Barathon
Journaliste

Projets

Chez Chronopost, la

technologie traite de la data pour le business



Hugues de Maussion, DSI de Chronopost, a été le Grand Témoin de la conférence Data Analytics organisée par CIO le 25 avril 2017.

Grand Témoin de la conférence « Data Analytics » organisée par CIO le 25 avril 2017, Hugues de Maussion, DSI de Chronopost, a détaillé ses emplois de l'IoT et de la Data Analytics au bénéfice de l'activité de son entreprise.

A la fin de l'année dernière, Hugues de Maussion, DSI de Chronopost s'était réjoui : « [IoT ou Big Data, certains en parlent, nous, nous le faisons.](#) » Ce DSI est revenu sur l'ensemble de ses projets autour de la data au service du business en étant le Grand Témoin de la Matinée Stratégique « Data Analytics » organisée par CIO le 25 avril 2017 à Paris.

Chronopost est l'entité en France de DPD Group (Geopost), une filiale du groupe La Poste. Avec plus de 6 milliards d'euros de chiffre d'affaires au niveau mondial (dont un milliard en France), cette société est spécialisée dans le transport de « petits » colis (moins de 30 kg) express en France et à l'international. Son position premium l'oblige à des services de haute qualité autant sur le marché B2B que B2C. « Notre métier est de livrer en J+1, voire en J, et le e-commerce contribue évidemment à notre croissance » a expliqué Hugues de Maussion. Il a observé une évolution nette du marché du transport en France qui est passé du transport de palettes sans vrai délai garanti à du petit colis à J+1, coeur aujourd'hui du marché. Surtout, de nouvelles offres de Chronopost, comme Chronofresh et Biologistic, reposent sur les avancées techniques autour du traitement de la donnée. Ces deux offres consistent à transporter en urgence des marchandises sensibles en température dirigée, dans l'alimentaire pour la première et le médical pour la seconde.

Le Big Data est le socle des ruptures technologies

Hugues de Maussion a constaté : « la majorité des industries est percutée par un grand nombre de nouvelles technologies. Pour Chronopost, c'est l'IoT, l'IA [Intelligence Artificielle], le conversationnel, la réalité augmentée, la robotique et même les véhicules autonomes. Toutes ont comme socle commun le Big Data. » Si certaines technologies ont du mal, souvent, à se généraliser malgré un discours ambiant très favorable,

celles-ci peuvent avoir un vrai rôle quasiment Legacy dans certains secteurs. « L'IoT, par exemple, nous sommes tombés dedans quand nous étions petits car, depuis toujours, nous menons du tracking de colis » a ainsi observé Hugues de Maussion. Le tracking de colis s'est enrichi avec le géo-positionnement depuis l'arrivée du GPS (autre objet connecté). Et, plus récemment, viennent de s'ajouter, pour les colis transférés en température dirigée, des capteurs de température pour ajouter cette information. De la même façon, l'IA permet d'améliorer la prise de décision. Le conversationnel, ce sont les bots qui améliorent l'expérience client en diminuant les coûts de l'interaction. « Beaucoup d'informations réclamées lors d'appels aux services clients pourraient être obtenues avec le tracking » a déploré Hugues de Maussion. Les bots facilitent ainsi l'obtention d'une information simple. Hugues de Maussion a relevé : « par contre, quand il y a un problème complexe, un humain doit reprendre la main. »

Des technologies parfois encore expérimentales

La réalité augmentée a un usage plus interne. Il s'agit en effet de permettre à des opérateurs de lire des informations dans des lunettes genre Google Glasses tout en manipulant des objets. Mais, chez Chronopost, des projets de ce genre sont restés expérimentaux. Autre sujet, la robotique. Celle-ci résulte d'un arbitrage entre CapEx/OpEx, selon un investissement dans un robot ou un coût récurrent en salaire. Le coût des robots diminue et, couplés à de l'intelligence artificielle, leur champ d'action s'étend.

Enfin, les véhicules autonomes sont aussi un sujet très actuel. Si beaucoup pensent aux Google Cars, le vrai sujet est plutôt du côté des drones. « Nous avons une ligne avec une liaison quotidienne pour livrer une technopole dans le Sud de la France » a pointé Hugues de Maussion. Il a concédé : « ce n'est certes pas cela qui constitue un chiffre d'affaires conséquent mais nous avons franchi toutes les étapes réglementaires et techniques. » Malgré tout, Chronopost travaille aussi sur des véhicules autonomes routiers. Ceux-ci ont des applications notamment au sein des entrepôts.

Un service premium avec délai garanti

Comme Chronopost est un service premium avec un délai d'acheminement garanti, à J+1 voire à J (commandé le matin, livré avant le soir), la chaîne logistique est volontairement désoptimisée. « Même si le camion n'est pas plein, il doit partir » a résumé Hugues de Maussion. L'échelle de valeur repose sur le temps. Traditionnellement, même à l'époque où le positionnement du colis n'était connu qu'avec des flashages de codes barres à chaque étape, le suivi était plus précis que chez d'autres transporteurs, avec beaucoup plus d'étapes de flashages, premium oblige.

La nouveauté est donc, avec Chronofresh et Biologistic, de rajouter un tracking de température. Hugues de Maussion a raconté : « nous avons regardé à l'étranger les tendances dans notre secteur et, au Japon, avec une population vieillissante qui se déplace moins, la livraison express à température dirigée était une tendance. » Avec les offres Chronofresh et Biologistic, dans chaque container contenant des colis frais (liste connue), une puce RFID permet le repérage et un capteur de température trace celle-ci. « Nous sommes les seuls transporteurs au monde à pouvoir donner la courbe de température d'un colis du point de collecte à la livraison » se réjouit Hugues de Maussion.

La révolution arrive aujourd'hui

Pour l'instant, ces services sont encore limités, « quelques milliers de colis par jour » du

propre aveu du DSI qui a anticipé cependant : « nous pensons qu'une offre comme celle-ci va révolutionner le marché de l'alimentaire frais, permettant d'être livré à partir d'un petit producteur local à qui on a commandé en ligne. » Evidemment, le service, plutôt onéreux, est destiné aux productions haut de gamme, en général de traditions régionales, mais il n'est pas dit que cela ne s'étende pas. En effet, ce type de livraison rend inutile une chaîne d'intermédiaires, chaque maillon prenant sa marge. Le coût supplémentaire du transport pourrait donc être compensé par la disparition de ces intermédiaires.

Pour l'heure, les capteurs de température sont relevés lorsque les containers bénéficient d'une connexion, dans les camions (4G) ou dans les locaux (Wi-Fi). Sur le dernier kilomètre, les données sont certes collectées mais elles ne sont transmises qu'au retour du container. Cependant, bientôt, les terminaux des livreurs du dernier kilomètre vont être remplacés et les nouveaux seront nativement connectés. Dès lors, la remontée d'informations sera en permanence temps réel.

La bonne technologie pour chaque cas

Pour traiter les données ainsi remontées, du vrai Big Data, plusieurs technologies sont employées. Pour commencer, Chronopost utilise deux types de bases de données NoSQL, Cassandra et Hadoop. Les deux sont en production depuis plusieurs années sur des clusters de chaque technologie : chacune a ses avantages et chacune doit être utilisée à bon escient. « Sur Cassandra, nous traitons le transactionnel massif temps réel, par exemple le déplacement du colis sur une carte » a spécifié Hugues de Maussion. Cassandra est en voie de remplacer progressivement toutes les autres bases de données transactionnelles chez Chronopost.

Par contre, Cassandra n'est pas approprié pour l'analytique, qui repose sur un cluster Hadoop. Les données du cluster Cassandra sont donc déversées dans un second cluster sous Hadoop à cette fin. Hugues de Maussion a précisé : « nous ne procédons pas à du streaming temps réel, qui n'aurait pas eu d'utilité mais un coût, le transfert est donc réalisé toutes les heures. » Différents outils de reporting sont placés par dessus les noeuds : Dataiku, SAP BO, outil maison... Cela dit, les volumes de données ne sont pas considérables, chaque donnée étant petite. Mais l'approche Big Data a plusieurs avantages : faible coût, scalabilité et surtout résilience. « Dans une base de données traditionnelle, vous avez une écriture unique dans une base tandis qu'une approche en cluster de noeuds implique une réplication native » a relevé le DSI. Si un noeud tombe, le service peut continuer, ce qui est essentiel pour les clients de l'entreprise.

Cassandra et Hadoop : deux choix de stockage différents

En dessous des bases de données, le stockage est une brique de base pour garantir le bon fonctionnement des services de Chronopost. « Il y a trois ans, avec nos attentes de niveaux de service, nous étions un peu seul face aux fournisseurs » a soupiré Hugues de Maussion. Le cluster Cassandra repose sur des serveurs-galettes bas de gamme, recyclés, avec stockage intégré, mais en grand nombre. Hugues de Maussion a expliqué : « chaque serveur Cassandra traite peu de données et cette architecture permet des temps de réponse extrêmement rapides, sans contention car, contrairement aux bases de données traditionnelles, l'écriture est répartie. »

Pour le cluster Hadoop, le choix technologique est différent pour le stockage. En effet, Chronopost a choisi un stockage Flash. « Je ne sais pas si je redirai la même chose dans cinq ans, mais pour l'instant, nous avons fait le choix de tout héberger chez nous car cela fait partie de notre coeur de métier » a confessé le DSI. Le recours à du cloud public a été écarté pour conserver la « souveraineté » des données.

En savoir plus

- [Contenus associés à la conférence Data Analytics du 25 avril 2017](#) (Progressivement enrichis).
- Télécharger l'étude [Quelles connaissances disponibles au-delà du Big Data ?](#)
- 28 Octobre 2016 : « [IoT ou Big Data, certains en parlent, nous, nous le faisons.](#) »

Bertrand Lemaire
Rédacteur en chef de CIO

Projets

Le prédictif au service du pilotage business



De la visualisation au prédictif » Godineau (Directrice Marque et ti (Product Director Engine & rcia De Soria (Directrice de la) ont témoigné sur « Le prédictif

(c) Bruno Levy

Il existe un domaine encore...
 f. Pourtant, de grandes...
 e business du prédictif.
 Bastien Albizzati (Product Director Engine & Marketplace chez Criteo) et Isabel Gomez Garcia De Soria (Directrice de la Recherche Opérationnelle chez Air France KLM) ont ainsi témoigné des projets menés dans leurs entreprises respectives en matière de prédictif au service du business. Ils se sont exprimés lors d'une table ronde intitulée « Le prédictif pour piloter l'activité » au cours de la Matinée Stratégique « Data Analytics : De la visualisation au prédictif » organisée par CIO le 25 avril 2017 à Paris.

Les projets présentés se déroulaient dans des contextes très différents. Ainsi, chez Air France KLM, il s'agissait d'IoT autour de très gros objets, en l'occurrence des avions A 380. Exploitant des liaisons en bus interurbaines, Ouibus a cherché à démontrer la pertinence business de chaque spot publicitaire diffusé à la télévision. La publicité est d'ailleurs le coeur d'activité de Criteo dont le métier est d'acheter des affichages de bandeaux publicitaires sur des sites web et de vendre aux annonceurs des clics sur leurs bandeaux, ce qui implique d'afficher un bandeau qui sera cliqué par l'internaute.

Du prédictif pour de la maintenance préventive

« Notre problème avec l'A380, un avion jeune, c'était un taux de panne plus élevé qu'avec des modèles d'avions plus matures » a expliqué Isabel Gomez Garcia De Soria. Or, à chaque panne, c'est un vol avec 500 passagers qui peut être annulé. Et certaines pannes peuvent survenir très loin d'un aéroport proposant des possibilités suffisantes de maintenance. Et la démarche engagée visait donc à réaliser une maintenance préventive pour faire en sorte que la panne n'ait pas lieu.

Or, chez Air France KLM, réaliser des prédictions à partir de Big Data est une activité courante depuis plusieurs années, notamment en termes de trafic ou de besoins clients. Les spécialistes du domaine se sont donc associés à des ingénieurs en avionique pour analyser les pannes, en particulier les plus fréquentes, dans les pompes à carburant. Isabel Gomez Garcia De Soria s'est réjouie : « l'avantage d'un avion jeune comme l'A380 est qu'il produit énormément de données que l'on peut exploiter. » Plusieurs giga-octets de données sont ainsi collectées à chaque fois qu'un avion arrive à Roissy Charles-de-Gaulle à partir des nombreux capteurs situés dans l'appareil, environ 1,6 Go/vol. Ces données étaient simplement stockées et non-exploitées.

Décider en quelques minutes

Ce sont sur celles-ci que des algorithmes prédictifs ont pu être appliqués. « Ces données ont été analysées en les rapprochant des pannes constatées dans telles ou telles conditions » a confié Isabel Gomez Garcia De Soria. Des modèles de pannes ont ainsi été définis. Les pièces en cause ont été changées en amont d'une défaillance et, ensuite, été testées avec des tests d'efforts, provoquant effectivement les pannes attendues, validant ainsi le modèle algorithmique.

A chaque déversement de données, les datas sont collectées dans le datalake de la

compagnie et il faut, en moins de 25mn, dire si oui ou non une panne va survenir et sur quelle pièce. Au delà, les mécaniciens n'auraient plus le temps de procéder à un changement avant le redécollage. A chaque panne évitée, le coût économisé peut se chiffrer en centaines de milliers d'euros. Les évitements de coûts dépassent le million d'euros chaque année, sans oublier le gain de satisfaction client.

Lancer une marque avec peu de moyens

Ouibus avait d'autres préoccupations. Née de la Loi Macron, cette filiale de la SNCF devait se construire une notoriété à partir de zéro mais avec des moyens limités. En effet, le marché de Ouibus est low cost. Donc, il fallait faire un développement commercial important pour atteindre rapidement de gros volumes mais sans trop dépenser. « A ce jour, nous en sommes à 4 millions de passagers transportés sur 140 destinations » a précisé Béatrice Godineau.

Pour assurer ce lancement, il fallait garantir un très bon rendement des investissements publicitaires. Béatrice Godineau a insisté : « nous avons une culture ROIste chez Ouibus. A côté du Digital, le média le plus puissant pour toucher un grand volume de population est la télévision mais c'est un média coûteux ». L'objectif était donc de tracker, spot par spot, le business généré sur le site web de Ouibus, soit immédiatement soit en décalé. La solution Realitics permet de réaliser ce tracking : suite à tel spot, x ont été voir le site dans la foulée et y ont effectivement acheté un trajet, éventuellement plusieurs jours après. Béatrice Godineau a observé : « notre clientèle réagit en général très vite et nous avons confirmé les résultats de nos études clientèle en constatant des conversions en deux ou trois jours. »

Décider d'afficher la bonne publicité en quelques millisecondes

Décider vite, cela peut vouloir dire en quelques millisecondes. Chez Criteo, le retargeting consiste à diffuser des bannières publicitaires personnalisées qui vont inciter les internautes à cliquer dessus puis à achever la démarche commerciale. Il s'agit donc de prédire le comportement de l'internaute. « L'achat d'inventaire d'emplacements publicitaires sur Internet est réalisé impression par impression et payée au site mais Criteo est payé par ses propres clients uniquement en cas de conversion sur le site de e-commerce » a détaillé Bastien Albizzati. Il faut donc prédire la probabilité que chaque internaute convertisse l'affichage en achat pour que Criteo optimise son bénéfice.

Et ce quasi-instantanément. Bastien Albizzati a précisé : « entre le moment où un affichage est proposé et celui où l'on doit proposer une bannière, il s'écoule moins de 20ms. » Il faut donc que Criteo soit très efficace, très pertinent et très rapide pour gagner de l'argent.

« Bien sûr, pour y parvenir, il faut traiter de la donnée, qui est, pour nous, le tracking de cookies anonymes » a explicité Bastien Albizzati. Si Criteo peut savoir que monsieur X a surfé sur plusieurs sites de voyages, l'identité de Monsieur X ne l'intéresse pas. Les algorithmes de machine learning travaillent en général sur un mois de données.

Retravailler les algorithmes

Les différentes sortes de prédictions nécessitent cependant un travail pointu sur les algorithmes. « Chez Air France, nous travaillons sur du prédictif dans plusieurs domaines -IoT ou marketing par exemples- et à chaque fois les algorithmes à appliquer sont spécifiques aux données traitées, même si les bases sont globalement les mêmes » a spécifié Isabel Gomez Garcia De Soria. Ensuite, il faut tester. Et améliorer en continu. Bastien Albizzati a confirmé : « il n'existe aucun algorithme magique

capable de prédire tout et n'importe quoi. »

En savoir plus

- [Contenus associés à la conférence Data Analytics du 25 avril 2017](#) (Progressivement enrichis).
- Télécharger l'étude [Quelles connaissances disponibles au-delà du Big Data ?](#)
- 7 Octobre 2016 : [Criteo prédit l'avenir avec une obligation de performance](#)
- 6 Avril 2016 : [Ouibus mesure l'efficacité de ses publicités télévisuelles sur son e-commerce](#)

Bertrand Lemaire
Rédacteur en chef de CIO

Pour toute demande concernant CIO.focus :

contact-cio@it-news-info.com

Une publication de IT NEWS INFO : 40 bd Henri Sellier 92150 Suresnes

Rédacteur en chef : Bertrand Lemaire, blemaire@it-news-info.com

Tél. : 01 41 97 62 10

Principaux associés : Adthink Media et International Data Group Inc.

Président : Bertrand Gros

Data Analytics : de la visualisation au prédictif

n°139 - Mai 2017

Directeur de publication : Bertrand Gros

Directeur général : Jean Royné

Président du groupe Adthink Media : Sylvain Morel

CIO est édité par IT NEWS INFO, SAS au capital de 3000000 €

Siret : 500034574 00029 RCS Nanterre

