



# Hortonworks®

## Le secteur de la vente au détail dope ses ventes avec Hadoop®

Une architecture de données moderne pour les détaillants

Un document de présentation  
technique de Hortonworks  
September 2014

# Sommaire

---

<b>Introduction</b>		<b>3</b>
<b>Marketing</b>	Profilage complet des clients	<b>6</b>
	Segmentation de la clientèle	<b>7</b>
	Transparence des prix	<b>8</b>
	Analyse de sentiment sur marque	<b>9</b>
<b>E-commerce et service client</b>	Moteur de recommandation de produit	<b>10</b>
	Optimisation du chemin d'accès Web	<b>11</b>
<b>Expérience en magasin</b>	Placement du produit	<b>12</b>
	Offres personnalisées en magasin	<b>13</b>
<b>Chaîne d'approvisionnement et logistique</b>	Gestion des stocks	<b>14</b>
	Préparation efficace des commandes	<b>15</b>
	Gestion des fournisseurs	<b>16</b>
	Approvisionnement stratégique	<b>17</b>
<b>Enterprise Hadoop</b>		<b>18</b>

---

## Introduction

Les effets de l'innovation technologique, de la modification des préférences des consommateurs et d'une période prolongée de ralentissement économique s'ajoutent pour provoquer des changements rapides dans le secteur de la vente au détail. Pour les détaillants, Apache Hadoop peut être un outil puissant de réponse à ces changements.

Dans beaucoup de pays, les taux de croissance des dépenses des consommateurs observés après la récession et considérés désormais comme la « nouvelle normalité » sont inférieurs à ce qu'ils étaient avant la crise. Dans le même temps, la technologie a permis aux consommateurs de devenir plus avertis et d'échanger davantage. Les évaluations, les analyses et les comparaisons de prix en ligne influencent leurs achats et, lorsqu'ils se décident, ils peuvent acheter sur un appareil mobile, en ligne ou dans un magasin. On peut désormais faire ses achats 24h/24, 7j/7.

Les ventes en ligne (e-commerce) continuent de progresser rapidement -15,5 % aux États-Unis en 2014 - et représentent désormais 6 % du montant total des ventes au détail.<sup>1</sup> Dans le même temps, les magasins physiques voient une croissance faible du marché ; ainsi, au final, la croissance globale des ventes au détail se situe en dessous de 4 % par an.<sup>2</sup> Ces tendances font que certains types de magasins, comme les agences de voyages ou les librairies, sont en train de disparaître. Les détaillants encore présents subissent une pression sur leurs marges face aux nouveaux acteurs du Web et aux attentes des consommateurs sur une plus grande transparence tarifaire.

Dans cet environnement, les détaillants ont besoin, en se basant sur les données, de segmenter les consommateurs, d'interagir avec eux et de gérer leur chaîne logistique pour faire correspondre leurs produits à la demande des clients.

Hadoop améliore la segmentation des clients. Les détaillants peuvent exploiter un nombre considérable de données clients provenant de sources jusque-là inaccessibles à grande échelle : les analyses en ligne, les blogs et les messages des réseaux sociaux. Les nouvelles sources de données de Hadoop fournissent aux détaillants des informations inédites sur les comportements et les préférences des consommateurs. Les données de Hadoop améliorent aussi les échanges avec les clients en permettant de leur faire des recommandations d'achat sûres. Elle permet aussi de reconnaître les clients à travers de multiples canaux et points de vente pour proposer des recommandations personnalisées dans le magasin ou en ligne, ainsi que des bons d'achat en temps réel pour aiguïser leur appétit de consommateur. Selon une étude de McKinsey, ces programmes de ventes au détail basés sur l'analyse de gros volumes de données peuvent accroître les marges des détaillants de 60 %.<sup>2</sup>

Enfin, les détaillants utilisent Hadoop pour améliorer leurs prévisions, leurs achats et leur chaîne logistique. Ils peuvent optimiser les circuits de distribution et les niveaux de stocks pour prendre en compte les goûts locaux et en évolution constante de chaque consommateur.

Les concepteurs-fondateurs de Hortonworks ont lancé Apache Hadoop YARN, qui a fait évoluer la plate-forme par rapport à sa stricte fonction batch d'origine. Hortonworks a investi fortement dans YARN et a conduit son développement au sein de la communauté Apache.

<sup>1</sup> Le total des ventes au détail aux États-Unis a dépassé 4500 milliards de dollars en 2013 et progresse plus vite que le PNB, eMarketer, 10 avril 2014 <http://www.emarketer.com/Article/Total-US-Retail-Sales-Top-3645-Trillion-2013-Outpace-GDP-Growth/1010756>

<sup>2</sup> Big data: The next frontier for innovation, competition and productivity, McKinsey Global Institute, juin 2011

## RÉSULTATS AVEC HADOOP

Le système de données de Luminar collecte

### 15 téraoctets provenant de 2000 sources brutes

Au lieu de 2 téraoctets provenant de 300 sources initialement

TrueCar stocke et traite des données au prix de

### 0,23 \$ par gigaoctet

Au lieu de 19 \$ par gigaoctet sur une plate-forme traditionnelle

Un grand magasin assure sa promotion avec

### Des livraisons en temps réel

Sur l'ensemble des canaux médias

L'analyse Clickstream pour les détaillants permet désormais

### de stocker à la fois les fichiers de parcours sur la toile et les réactions aux publicités

Dans le même data lake

Désormais, Hadoop 2 (avec YARN) intègre les capacités renforcées suivantes, que personne ne peut mieux connaître que Hortonworks.

**Traitement des données multiutilisation, multicharge :** Hadoop supporte de nombreuses méthodes d'accès (batch, interactif, temps réel) à un ensemble commun de données. Les analystes peuvent visualiser et transformer instantanément les données de différentes façons. Cela accélère l'accès aux informations et rend ces analystes plus confiants sur leurs résultats.

**Nouvelles opportunités pour l'analytique :** L'architecture de schéma à la lecture de Hadoop permet aux utilisateurs de stocker les données sous leur format brut. Les analystes définissent ensuite les schémas uniques pour les données dont ils ont besoin pour une question ou une application particulière.

**Des gains d'efficacité pour l'architecture des données :** Hadoop fonctionne sur des serveurs standard à faible coût et réduit le coût global du stockage. Cela permet de conserver, à un coût abordable, toutes les données sources pendant de plus longues périodes, ce qui autorise des applications avec un historique beaucoup plus fourni.

**Optimisation de l'entrepôt de données :** Les charges de travail ETL bénéficient aussi du coût favorable de Hadoop. Les données, dont la valeur unitaire est faible, peuvent être extraites et transformées dans Hadoop. La valeur de ces données augmente dans l'agrégat, lorsque d'autres données viennent s'y rajouter et que le temps de stockage se prolonge.

Ces avantages expliquent pourquoi les détaillants ont adopté Hadoop partout dans le monde. De fait, la plateforme de données Hortonworks (HDP) est actuellement utilisée par la plupart des grandes sociétés de vente au détail aux États-Unis et en Europe.

Ce document de présentation technique illustre, à travers douze cas d'utilisation classiques, les améliorations pratiques apportées dans les entreprises par Hadoop. Les détaillants tirent le meilleur bénéfice de Hadoop lorsqu'ils se concentrent d'abord sur les problèmes spécifiques à leur secteur d'activité, comme ceux détaillés ci-dessous. À cela s'ajoute la valeur que Hadoop crée dans les fonctions horizontales comme l'informatique, mais celles-ci ne sont pas spécifiquement couvertes dans le document présent.

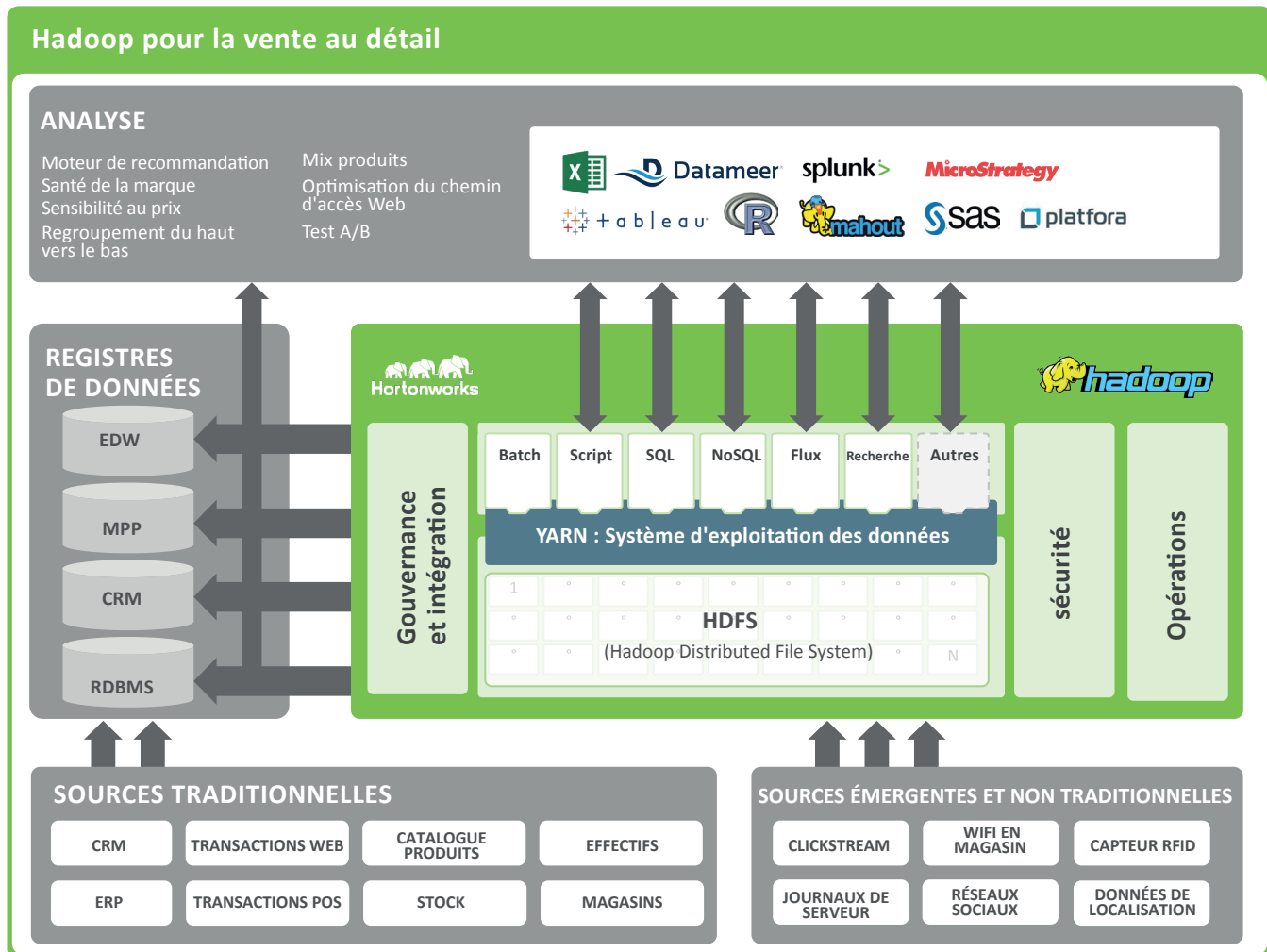


Illustration 1: Ingérer toutes les données, les stocker à grande échelle et l'analyser avec les outils que vous utilisez déjà.

## Marketing

Hadoop permet de doper la performance marketing grâce à sa capacité à améliorer la segmentation, la conversion, la part de portefeuille et la valeur à vie des clients.

### VISION COMPLÈTE DE LA VALEUR À VIE DES CLIENTS

#### Les défis pour l'entreprise

Les détaillants interagissent avec les clients par de nombreux canaux, c'est pourquoi les données d'échange et de transaction avec les clients sont souvent stockées dans différents silos de données. De ce fait, peu de détaillants peuvent corréler précisément les achats de leurs clients avec des campagnes de marketing et leur comportement de navigation en ligne afin de pouvoir prédire la valeur à vie de ces clients (VVC).

Pour donner un exemple : un détaillant important de produits d'amélioration résidentielle vend en ligne et dans des milliers de magasins dans le monde. Mais les données sur le trafic de son site Web, les transactions POS et les services à domicile n'existaient que dans des silos fragmentés. À défaut d'un « enregistrement en or » unifié des comportements des clients, le détaillant ne pouvait quantifier que difficilement la valeur à vie de ses échanges. En l'absence d'informations détaillées de segmentation, l'entreprise ne pouvait pas prioriser ses efforts de marketing en fonction de la VVC attendue.

#### Solution

Le détaillant a installé Hadoop à son siège social pour consolider toutes les informations clients dans un seul data lake. Désormais, cet « enregistrement en or » permet de relier les données structurées ERP et CRM avec les données du clickstream, des réseaux sociaux, de géolocalisation et des transactions POS. L'entreprise utilise ce data lake pour obtenir une segmentation précise, très ciblée et des calculs de VVC. Ces informations complètes conduisent à des actions spécifiques en direction du public, comme des bons d'achat personnalisés, des promotions et des e-mails.

#### Impact

À présent, Apache Hadoop fournit à ce détaillant un profilage complet de ses clients, segmenté selon plusieurs domaines, notamment l'analyse des paniers d'achats, les préférences exprimées sur les réseaux sociaux et les retours observés sur les campagnes de marketing. Une meilleure analytique augmente les ventes, réduit les frais de stock et améliore la valeur à vie des clients. Chaque année, les économies d'échelle permises par Hadoop font gagner à l'entreprise plusieurs millions de dollars sur ses dépenses récurrentes dans les entrepôts de données.

## SEGMENTATION DE LA CLIENTÈLE

### Les défis pour l'entreprise

Luminar, une division d'Entravision Communications Corporation, est la société leader en analyses de big data ; son activité principale consiste à fournir des données exploitables sur les consommateurs latino-américains aux États-Unis. Les adultes latino-américains disposent d'un pouvoir d'achat annuel de 1500 milliards de dollars et les publicitaires utilisent l'analyse de Luminar pour toucher ce public.

Le problème de Luminar était de pouvoir stocker suffisamment de données empiriques pour fournir des recommandations fiables à ses clients détaillants. Le fait d'utiliser les mêmes données d'échantillonnage, incomplètes et obsolètes, que les autres organes de recherche marketing ne lui convenait pas. Les données de Luminar provenaient de 300 sources différentes qui concernaient environ 15 millions d'adultes latino-américains aux États-Unis, notamment les transactions sur les cartes de crédit, les données des décodeurs de télévision, les données de vote et les réseaux sociaux. L'entreprise avait besoin d'une plate-forme stable pour stocker et traiter ces données complexes.

### Solution

Luminar a choisi Apache Hadoop pour combler ses exigences en matière de coût, d'efficacité et de flexibilité. Hadoop stocke et traite les données qui alimentent les modèles de ciblage de Luminar, et l'entreprise sait que son environnement Hadoop peut évoluer avec son activité. Désormais, Luminar peut gérer simultanément un plus grand nombre de projets et de données de traitement pour davantage de clients. De plus, elle peut maintenant isoler les signaux marketing uniques et exploitables liés aux besoins des clients.

### Impact

Luminar a multiplié par huit le nombre de données enregistrées. Auparavant, elle enregistrait tous les mois 2 téraoctets de nouvelles données transactionnelles à partir de 300 sources brutes. Avec la plate-forme de données Hortonworks, elle acquiert désormais 15 téraoctets provenant de 2000 sources brutes. Avant Hadoop, il fallait plusieurs jours à Luminar pour enregistrer les données et y accéder avant de rafraîchir un modèle. Désormais, cela ne prend que quelques heures. Luminar propose à ses meilleurs clients un portail d'exploration de données en self-service à travers l'interface Web Tableau.

## TRANSPARENCE DES PRIX

### Les défis pour l'entreprise

La plupart des acheteurs de voitures peuvent consulter en ligne les marques et les modèles avant d'aller chez un concessionnaire local pour obtenir la dernière information critique : le prix. Le manque de transparence sur les prix alimente la méfiance des clients et ralentit le processus de vente. Cela nuit à la fois aux acheteurs et aux vendeurs. Lorsqu'on les interroge, les clients croient que les distributeurs réalisent une marge moyenne de 19,7 % et jugent que 13,2 % serait un chiffre raisonnable.<sup>3</sup> En réalité, la marge moyenne d'un distributeur est plus proche de 3 à 4 %.<sup>4</sup>

La mission de TrueCar est de fournir en toute transparence des informations justes pour accroître la confiance et améliorer l'expérience d'achat d'une voiture. L'entreprise acquiert et analyse des quantités phénoménales de données afin d'estimer ce que les autres acheteurs ont payé pour un véhicule donné. Comme le nombre de données et la complexité de l'analyse augmentaient, l'entrepôt de données sur le serveur SQL existant de TrueCar devenait de plus en plus coûteux et difficile à étendre.

### Solution

TrueCar a transféré l'ensemble de ses données de prix critiques sur un data lake Hadoop. Grâce à Apache Flume, les données des journaux du serveur provenant de plus de 1000 sources se déversent dans Hadoop. Les données recouvrent environ 8400 distributeurs, 8 millions de véhicules et 250 millions de photos de voitures.

TrueCar enrichit ces données via des centaines de processus et propose des estimations de prix locales avec un degré élevé de confiance sur ses applications mobiles, ses plus de 400 sites partenaires et son site vedette TrueCar.com. L'entreprise a facilité la vente de plus de 1,4 million de voitures à ce jour.

### Impact

TrueCar a bénéficié de deux avantages substantiels. D'abord, elle peut mieux servir ses clients (acheteurs et distributeurs) en stockant et traitant suffisamment de données pour fournir des estimations de prix fiables, son système pouvant de surcroît être étendu au fur et à mesure que davantage de voitures et de distributeurs y sont intégrés. Ensuite, chose tout aussi importante, l'entreprise stocke et traite désormais ses données au prix de 0,21 € (0,23 USD) le gigabit. Une plate-forme traditionnelle coûterait 17 € (19 USD) par gigabit, sans aucune capacité de traitement.

3) Version TrueCar de l'étude 2013 Strategic Vision New Vehicle Experience Study

4) National Automobile Dealers Association (Association nationale des distributeurs automobiles)

## ANALYSE DE SENTIMENT SUR UNE MARQUE

### Les défis pour l'entreprise

Les entreprises ne disposent pas d'un moyen rapide et fiable pour suivre la santé de leur marque. Elles peuvent avoir du mal à analyser la façon dont la publicité, les décisions des concurrents, les lancements de produits ou les nouveaux magasins affectent la perception de leur marque par les clients. Les études de marque ou les groupes cibles traditionnels prennent du temps et sont coûteux et sensibles aux biais d'échantillonnage.

Aujourd'hui, l'analyse de sentiment sur les réseaux sociaux peut produire un volume élevé d'informations tangibles et ciblées sur le ressenti des clients. Cela nécessite une plate-forme flexible et économique pour capturer, stocker et traiter ces flux de données sociales. Les bases de données traditionnelles du type « schéma au chargement » ne sont pas adaptées au contenu très variable des posts sur les réseaux sociaux et leur coût relativement élevé limite le nombre de cas d'utilisation pratiques. Les détaillants doivent choisir un nombre limité d'hypothèses avant d'explorer les données.

### Solution

Hadoop permet de saisir rapidement et sans biais les opinions exprimées sur les marques dans les réseaux sociaux. Les analystes peuvent explorer de façon interactive les avis sur Twitter, Facebook, LinkedIn ou sur les flux de réseaux sociaux spécifiques à un secteur d'activité. Les modifications des opinions des clients peuvent ainsi être capturées tous les jours (ou toutes les heures). Les analystes peuvent aussi rechercher des mots ou des phrases couramment utilisés par les défenseurs et les détracteurs de l'entreprise, en vue de les utiliser pour de futures campagnes.

### Impact

En appréhendant mieux les perceptions des clients, les détaillants peuvent mieux faire correspondre leur communication, leurs produits et leurs promotions à celles-ci. Ils peuvent réaliser des microévaluations sur le retour d'investissement de leurs efforts commerciaux, en mesurant les sentiments positifs sur la marque exprimés par des segments spécifiques de clientèle, dans des lieux particuliers et à différents moments de l'année. Avec l'architecture de schéma à la lecture de Hadoop, le détaillant peut explorer instantanément de nombreuses hypothèses (et en ajouter d'autres au fur et à mesure de son exploration).

## E-commerce et service client

Hadoop permet de doper la performance marketing grâce à sa capacité à améliorer la segmentation, la conversion, la part de portefeuille et la valeur à vie des clients

### MOTEUR DE RECOMMANDATION DE PRODUIT

#### Les défis pour l'entreprise

Un grand département spécialisé de taille importante voulait améliorer le ciblage de ses campagnes de marketing. L'équipe de marketing voulait lancer des promotions, des bons d'achat et des recommandations de produits personnalisés sur plusieurs points de contact clients : magasins, kiosques, Web et application mobiles. L'entreprise souhaitait en particulier mettre en place en temps réel des promotions sur les produits en magasin.

Mais les données clients de l'entreprise étaient fragmentées et cela l'empêchait de lancer des promotions commerciales sur la base de données. Par exemple, le site Web ou un kiosque ne recommanderait pas le produit qu'un client aurait déjà acheté dans le magasin. Les obstacles financiers et politiques bloquaient toutes les propositions de projets informatiques visant à moderniser l'architecture des données.

#### Solution

À présent, un data lake Hadoop intègre toutes les données brutes des clients sur différentes gammes de produits. L'entreprise acquiert les données en temps réel et par batch, à la fois sous des formats structurés et non structurés. Un processus ETL utilise Apache Pig et Apache Hive pour transformer les données brutes, qui sont ensuite digérées par les algorithmes d'apprentissage automatique Mahout and R. Notre détaillant peut maintenant proposer des recommandations et des promotions en temps réel à travers tous ses canaux, notamment son site Web, ses kiosques de magasin et ses applications mobiles.

#### Impact

Ce détaillant a bâti un moteur de recommandation multicanal similaire au moteur en ligne d'Amazon. Trente-cinq pour cent de ce que les clients achètent sur Amazon et 75 % de ce qu'ils regardent sur Netflix sont liés à des recommandations de produits basées sur ce type d'analyse.<sup>5</sup> Ce détaillant peut modifier ses recommandations en fonction de la météo, de la fidélité du client, de l'historique d'achat, des paniers abandonnés ou d'événements particuliers de la vie – et les produire à ses clients dans ses magasins

5) *Big Data and the Creative Destruction of Today's Business Models*, A.T. Kearney, 2013, <http://www.atkearney.com/documents/10192/698536/Big+Data+and+the+Creative+Destruction+of+Today's+Business+Models.pdf>

## OPTIMISATION DU CHEMIN D'ACCÈS WEB

### Les défis pour l'entreprise

Les acheteurs en ligne laissent des milliards de données lors de leur parcours sur la toile (clickstream). Les données du clickstream indiquent aux détaillants les pages Web que les consommateurs ont visitées et ce qu'ils ont acheté. De façon encore plus importante, le clickstream donne des indices sur les raisons de l'échec d'une transaction en ligne. Ceci ne peut cependant être fait qu'à grande échelle (dix parcours sur la toile ne constituent pas un modèle).

Mais le volume gigantesque de connexions non structurées sur le Web est difficile à appréhender, stocker, affiner et analyser pour avoir des informations. Le stockage des données de journaux Web dans des bases relationnelles nécessite des transformations complexes pour les convertir en un schéma exploitable. C'est également très coûteux. La valeur marginale de chaque nouveau parcours sur la toile ne justifie pas toujours son coût de stockage, aussi les négociants en ligne font des hypothèses et choisissent des ensembles de données dont ils attendent qu'ils confirment ou infirment lesdites hypothèses.

### Solution

Les détaillants du Web utilisent les données du clickstream pour comprendre les parcours des utilisateurs, analyser les paniers, effectuer des tests A/B et fixer des priorités pour les mises à jour de leur site. Apache Hadoop peut stocker tous les journaux Web pendant des années et à un prix faible. La flexibilité et l'efficacité de Hadoop sur les volumes élevés permettent aux négociants en ligne de prévoir davantage de tests simultanés sur les mêmes données brutes. Les outils d'accès aux données comme Hive, HBase, Storm et Solr permettent aux analystes d'effectuer des analyses récurrentes ou au coup par coup.

### Impact

Un client de Hortonworks réalise une analyse du clickstream pour l'un des plus importants détaillants au monde. Les données du clickstream étaient enregistrées au rythme de plusieurs centaines de mégabits par heure et de milliards de lignes par mois. À présent, l'agence stocke les fichiers d'impression sur les publicités et les fichiers de parcours dans le même data lake, et peut facilement y accéder par la suite pour des informations sur les clients. Avec un meilleur ciblage, leur client a réduit le nombre de publicités – mais elles sont plus efficaces –, ce qui améliore à la fois l'expérience Web globale et ses ventes en ligne.

## Expérience en magasin

Les détaillants utilisent Hadoop pour stocker et traiter des données sur la manière dont leurs clients interagissent avec l'environnement des magasins. Ils utilisent ces informations pour améliorer l'affectation des espaces, présenter des produits plus efficacement et proposer des offres en temps réel dans les magasins.

### PLACEMENT DES PRODUITS ET DISPOSITION DES MAGASINS

#### Les défis pour l'entreprise

Un détaillant multicanal de taille importante savait que la disposition des magasins, le merchandising et le placement des produits avaient un impact sur les ventes. Malheureusement, les magasins physiques de l'entreprise n'avaient pas la vision « en amont de la caisse » de la façon dont les clients prenaient leurs décisions d'achat. L'entreprise voulait avoir le même niveau de visibilité et d'analyse sur le parcours physique des clients que celui donné par les données du clickstream pour les consultations de leur site Web.

Les capteurs de magasin, les étiquettes RFID et les codes QR pouvaient combler ce vide de données, mais ces technologies généraient des formats et des volumes de données que les anciens systèmes de l'entreprise ne pouvaient pas manipuler aisément. Le détaillant se convertit relativement tôt à Hadoop, car la plate-forme ne nécessitait pas un concept « schéma au chargement » qui aurait empêché l'acquisition et le stockage des données de localisation nécessaires pour le programme.

#### Solution

L'entreprise a commencé à utiliser les technologies iBeacon d'Apple dans ses principaux magasins. iBeacon capture les données de localisation au sein des magasins à partir des applications iPhone et Android des portables des clients. Les données sont ensuite introduites dans Hadoop et révèlent comment les clients se déplacent dans les magasins (ce qui peut alors être comparé aux emplacements de certaines catégories de produits). Au fur et à mesure que le programme iBeacon s'étend, Apache Hadoop peut stocker et traiter cet énorme volume de données provenant de capteurs et de la microlocalisation.

#### Impact

Les résultats de cette étude pilote spécifique – qui ne sont pas publics – montrent que le programme analytique big data du détaillant a fait bondir ses ventes de 10 %. Comme les données dans Hadoop permettent à l'entreprise d'optimiser sa connaissance des magasins, elles font apparaître des bénéfices supplémentaires, comme la réduction des stocks non nécessaires et l'amélioration de la satisfaction des clients grâce à un placement plus intelligent des produits et une refonte de la disposition des magasins.

## OFFRES PERSONNALISÉES EN MAGASIN

### Les défis pour l'entreprise

Un autre grand magasin spécialisé voulait personnaliser ses offres et les présenter en temps réel (plutôt que d'adopter l'approche uniforme généralement utilisée pour les promotions).

La fragmentation des données était le principal obstacle à la préparation de ces offres rapides et personnalisées. L'EDW contenait à la fois des données transactionnelles et non transactionnelles. Le coût relativement élevé de stockage et l'architecture « schéma au chargement » des systèmes originaux faisaient qu'il n'était pas envisageable, ni économiquement ni techniquement, de stocker toutes les données aussi longtemps que le détaillant l'aurait souhaité.

Le flux de données provenant de tous les canaux devait arriver sur un seul registre, avant que les offres ne puissent être déclinées en retour vers ces mêmes canaux : le Web, les kiosques en magasins et les applications mobiles.

### Solution

Cette nouvelle capacité, supportée par Hadoop, a été mise en œuvre par l'entreprise dans cinq de ses magasins. Les clients qui entraient dans l'un de ces magasins pouvaient accepter ces offres via leur smartphone. Si une cliente s'attardait au rayon chaussures, elle pouvait recevoir un bon d'achat pour des chaussures qu'elle avait regardées en ligne mais qu'elle n'avait pas achetées. Elle pouvait ensuite utiliser ce bon d'achat dans le magasin ou en ligne. Le détaillant est en train de vérifier l'hypothèse que les clients acceptent plus facilement des offres lorsqu'ils sont déjà dans le magasin.

### Impact

Les grossistes ont depuis longtemps amélioré leurs ventes avec des bons d'achat spécifiques à un produit imprimés sur le dos des reçus POS, mais ces bons sont limités et ne sont que des offres statiques proposées après l'achat. Désormais, ce détaillant stocke les données dans Hadoop et formule des recommandations de produits avec des outils comme Mahout and R. Il propose ensuite ses offres à travers tous les canaux avec une application comme Apache HBase. Ce système peut proposer plusieurs offres en temps réel et ces offres peuvent changer de manière dynamique en réponse aux variations des mix clients et produits.

## Chaîne d'approvisionnement et logistique

Hadoop améliore la performance de l'entreprise pour la gestion de ses achats et de sa chaîne logistique en accroissant l'efficacité sur l'ensemble de la chaîne de valeur. Cela favorise les décisions d'optimisation tactique et la réorientation stratégique de l'environnement fournisseurs.

### GESTION DES STOCKS

#### Les défis pour l'entreprise

Les marges des détaillants sont toujours aléatoires et une meilleure gestion des stocks contribue au maintien de prix qui dégagent du bénéfice. Par exemple, un battage publicitaire sur de nouveaux produits peut autoriser un prix de vente élevé – mais ces produits très demandés doivent être disponibles. Les détaillants qui sous-estiment une grosse demande peuvent manquer rapidement de ces produits très lucratifs et perdre des ventes.

Le surstockage est également très coûteux et conduit à des démarques qui réduisent les marges et la perception de valeur pour les clients. Certains détaillants y répondent en expédiant les stocks excédentaires dans un autre site, mais cela augmente les coûts et entame un peu plus la marge. Les détaillants doivent gérer leurs stocks en temps réel, en ayant une visibilité totale. Tout comme dans l'histoire de Boucle d'or, le détaillant prospère ne doit en avoir ni trop, ni trop peu, mais juste la bonne quantité.

#### Solution

Hadoop aide à surmonter ces difficultés en fournissant une vision en temps réel des ventes de produits, des indicateurs avancés sur la demande de produits, ainsi que des données plus nombreuses et plus approfondies sur la chaîne logistique. Cela aide à optimiser les niveaux de stocks et peut déclencher une reconstitution automatique de ceux-ci. Les données textuelles, du clickstream et des capteurs complètent les données structurées sur les historiques d'achat et les délais de livraison des produits. Ces données combinées concernent les entrepôts, les intermédiaires en gros, les partenaires logistiques, les outils de réseaux sociaux et les magasins.

#### Impact

Une étude d'A.T. Kearney<sup>6</sup> cite une chaîne de vente au détail qui a « transféré rapidement des centaines de millions de dollars de stocks excédentaires d'un magasin vers d'autres magasins et qui a dans le même temps établi un modèle prédictif de distribution pour limiter le surstockage. Cette chaîne collecte désormais les données sur les prix, les promotions et les cartes de fidélité pour savoir de manière encore plus détaillée ce qu'achètent les clients, quand et pourquoi. » L'entreprise suit à présent la totalité de son stock et non plus seulement les 100 premiers articles en surstockage.

6) *Big Data and the Creative Destruction of Today's Business Models*, A.T. Kearney, 2013, <http://www.atkearney.com/documents/10192/698536/Big+Data+and+the+Creative+Destruction+of+Today's+Business+Models.pdf>

## PRÉPARATION EFFICACE DES COMMANDES

### Les défis pour l'entreprise

La préparation des commandes dans un entrepôt est un processus très gourmand en main d'œuvre. Un centre de traitement peut effectuer des prélèvements par pièce, par zone, par lot, par boîte, ou une combinaison de ces méthodes. Mais il s'agit d'un processus reproductible sur de gros volumes, similaire à une ligne de montage. Une analyse en temps réel et historique des données de préparation de commande est nécessaire pour améliorer de manière continue ces opérations. Une préparation de commandes plus rapide entraîne des livraisons plus rapides, ce qui améliore le traitement des commandes et la satisfaction des clients

Les détaillants en ligne les plus importants, comme Amazon, utilisent un ensemble de mécanismes automatisés pour préparer les commandes. L'efficacité du stockage et du traitement dans Hadoop peut désormais aider aussi des détaillants plus petits qui veulent améliorer leurs processus de préparation de commandes (sans robots ni investissements très lourds).

### Solution

Hadoop stocke des données provenant de différentes sources, comme les commandes, le stock de produits, les configurations d'entrepôts et les capteurs de flux de pièces. Ceci peut être analysé sur l'ensemble du processus de prélèvement en fonction des priorités de l'entreprise, afin de pouvoir le dépanner et prendre des décisions sur les effectifs et les équipements. Avec suffisamment de données, des algorithmes d'apprentissage automatisé peuvent simuler le processus de préparation de commandes et proposer des optimisations.

### Impact

L'entreprise Western Digital, est un client de Hortonworks qui a optimisé son procédé de fabrication avec la HDP. Les entrepôts de traitement peuvent obtenir le même type de résultats : des réactions plus rapides face à des changements, des niveaux plus faibles de casses et de retours et des réductions de coût. Les modifications peuvent être simulées au moyen d'outils de BI fonctionnant avec les données Hadoop, avant leur déploiement sur le processus final dans les entrepôts et magasins.

7) <http://hortonworks.com/customer/western-digital>

## GESTION DES FOURNISSEURS

### Les défis pour l'entreprise

La plupart des détaillants travaillent avec de multiples fournisseurs dans leurs chaînes logistiques. Parmi ceux-ci figurent les intermédiaires en gros, les fournisseurs logistiques externes, les transporteurs et les fournisseurs d'emballage. Par manque de temps, il peut être difficile pour les détaillants d'analyser et de noter la qualité relative de leurs fournisseurs. L'analyse réalisée par Hadoop donne une vision objective de la performance des fournisseurs sur un ensemble d'indicateurs de performance clés (KPI). Ces KPI incluent la rentabilité du fournisseur, le respect de ses délais de livraison, les retours et les plaintes des clients.

De plus, les fournisseurs de produits génèrent de plus en plus de données numériques sur leurs offres et leurs niveaux de stocks et les rendent disponibles pour les détaillants. Les détaillants qui intègrent ces données dans leurs propres activités peuvent gérer plus efficacement les relations avec leurs fournisseurs. Par exemple, ils peuvent définir des alertes pour des expéditions à la demande sur des produits très vendus lorsque les données d'évolution des ventes prévoient une rupture de stock.

### Solution

Les détaillants intègrent dans Hadoop les données des fournisseurs en y ajoutant leurs propres données ERP, les informations des réseaux sociaux relatives aux livraisons des fournisseurs, les données POS et les quantités en stock. Comme beaucoup d'autres solutions data lake dans Hadoop, cette association de données est très utile pour beaucoup d'applications. Le traitement par batch en fonction de l'historique peut proposer des recommandations d'achat sur de nombreux fournisseurs et produits. L'exploration interactive des données SQL fournit des informations utiles lors des négociations avec les fournisseurs. L'analyse du streaming en temps réel peut générer des alertes si les KPI du fournisseur dépassent les limites prédéfinies.

### Impact

Une analyse standardisée des KPI des fournisseurs permet aux détaillants d'optimiser leurs relations avec ces derniers sans effort supplémentaire ni modification sur les opérations existantes. Les fournisseurs bénéficient aussi de cette transparence sur les données, car ils reçoivent ainsi des signaux clairs sur les attentes du détaillant et sur ce qu'ils doivent faire pour garder sa confiance. Par le fait, une plus grande disponibilité des données permet aux fournisseurs et aux détaillants de limiter leurs échanges téléphoniques – ce qui permet de se concentrer sur la livraison et la vente des produits.

## APPROVISIONNEMENT STRATÉGIQUE

### Les défis pour l'entreprise

Les détaillants très avancés adoptent une approche holistique pour optimiser la performance de leur base fournisseurs. Ils cherchent le moyen d'exploiter leur puissance d'achat à travers des stratégies comme la concentration des volumes, la transparence des prix de composants et la tarification variable ou indexée. Ces détaillants qui adoptent les meilleures pratiques visent aussi à générer des avantages technologiques par la modification des caractéristiques des produits et l'amélioration des procédés ou en réorganisant les relations au sein de la chaîne de valeur.

Habituellement, les détaillants utilisent les enchères pour leurs approvisionnements stratégiques, mais les données d'enchères sont complexes. Pour citer un exemple, un détaillant a utilisé les enchères pour des services logistiques externalisés et de retour. Plus de 100 transporteurs ont fait des offres pour tout ou partie des 10 000 itinéraires différents dont chacun avait plus de 20 caractéristiques, comme « fréquence » ou « contrôle de température ». Les systèmes traditionnels sont dépassés lorsque l'analyse s'étend aux textes non structurés, aux images et aux données GPS contenus dans les offres. Et lorsque la même enchère se répète, les données sur la nature et les résultats des enchères précédentes sont rarement disponibles.

### Solution

Hadoop fournit la plate-forme idéale pour combiner, corréler et interroger ces diverses sources de données. Cela permet au détaillant d'effectuer de nombreuses simulations simultanées pour obtenir les résultats des différentes combinaisons d'attribution de marché. Les fournisseurs sont encouragés à livrer des informations complexes, avec autant d'offres groupées ou conditionnelles et de conditions contractuelles créatives que possible. Dans certains cas, ces offres groupées s'étalent sur trois à cinq ans, concernent plusieurs groupes d'approvisionnement et pays et représentent plusieurs millions de dollars annuels. À défaut d'un système de données capable de s'adapter au nombre et à la complexité de ces informations issues des fournisseurs, la plupart d'entre elles ne peuvent plus être gérées par les décideurs. Avec Hadoop, plus de données signifie de meilleures décisions, ce qui accroît les profits.

### Impact

Les détaillants qui adoptent cette démarche d'approvisionnement stratégique, rigoureuse et riche en données, réalisent classiquement entre 10 et 15 % d'économie sur le prix des marchandises vendues. De plus, ils se constituent un actif de données internes pour analyser année après année comment améliorer leur processus d'approvisionnement. La période d'achat des fêtes commence tous les ans au même moment, mais les équipes d'acheteurs ne repartent pas de zéro à chaque fois. Avant de prendre des décisions, ils peuvent à présent « se souvenir » en détail des achats des années précédentes. Grâce à Hadoop.

---

Avec Hadoop, les détaillants peuvent prévoir les comportements d'achat des clients pour personnaliser les prix, les espaces et les assortiments dans les magasins. Ils peuvent améliorer leurs contacts avec les clients par tous les canaux : en ligne, par téléphone ou dans le magasin. Ils peuvent optimiser leurs approvisionnements internationaux, gagner en efficacité et réduire le prix des marchandises vendues.

Ces exemples de sociétés de vente au détail montrent ce que les entreprises peuvent apprendre des autres secteurs d'activité : Hadoop génère des économies substantielles par rapport aux anciennes technologies de stockage de données et crée de formidables nouvelles capacités pour faire croître l'activité.

## Bâtir une architecture de données moderne pour la vente au détail avec Enterprise Hadoop

Pour obtenir un bon retour sur votre investissement dans le big data, planifiez l'intégration d'Enterprise Hadoop avec votre EDW et les systèmes de données associés. En bâtissant une architecture de données moderne, votre entreprise pourra stocker et analyser de façon massive les données les plus importantes pour votre activité, extraire les informations critiques provenant de toutes les données et sources pour, en final, améliorer votre position concurrentielle sur le marché, fidéliser vos clients et maximiser votre chiffre d'affaires. Découvrez plus de détails sur : <http://hortonworks.com/hdp>

## La plate-forme de données Hortonworks supporte Enterprise Hadoop

La plate-forme de données Hortonworks (HDP) est administrée à 100 % par la solution open source Apache Hadoop. HDP fournit tous les projets liés à Apache Hadoop qui sont nécessaires à l'intégration de Hadoop en parallèle avec un EDW dans le cadre d'une architecture de données moderne. Elle est livrée avec une capacité puissante de gestion des données, une capacité polyvalente pour l'accès aux données, ainsi qu'avec trois capacités essentielles pour son adoption par les entreprises : Gouvernance et intégration des données, Sécurité et Opérations.

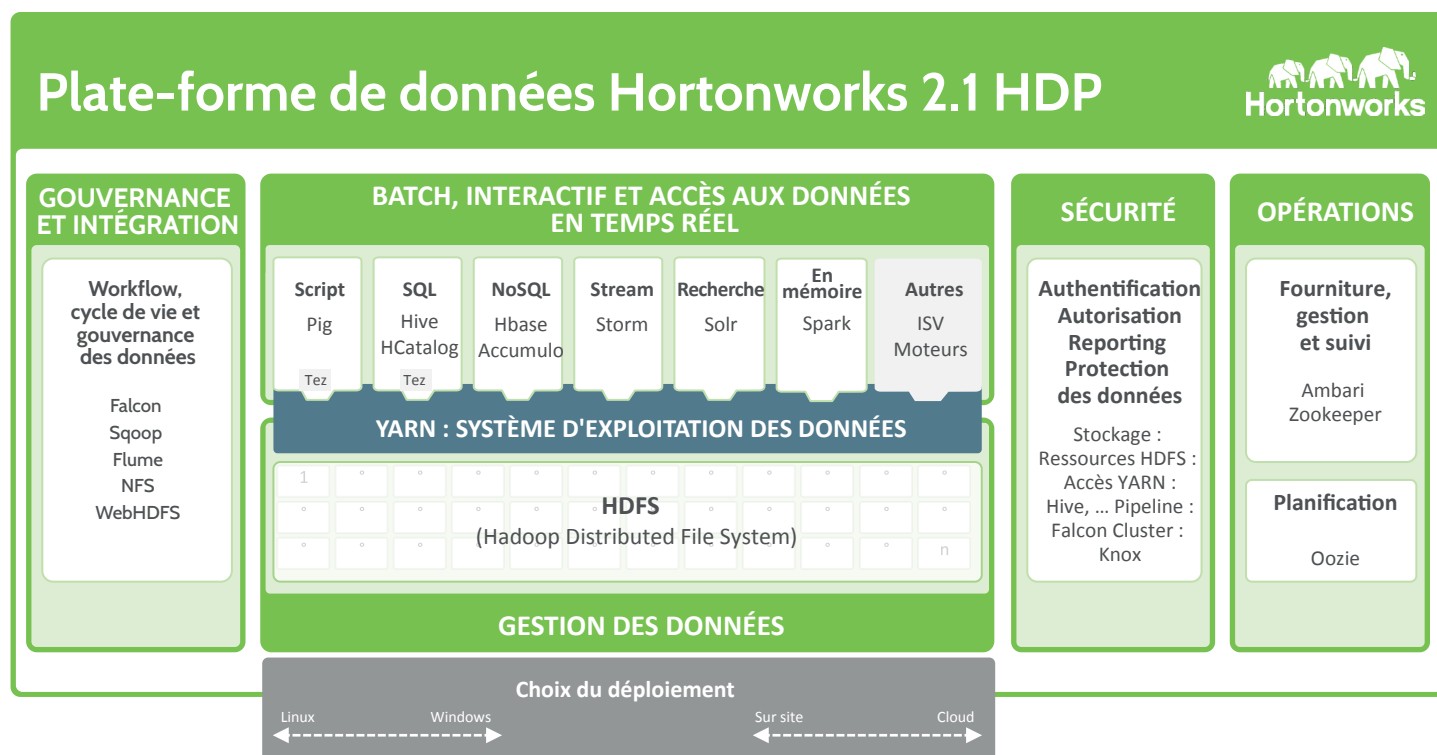


Illustration 2 : Cinq capacités centrales – gouvernance et intégration des données, gestion des données, accès aux données, sécurité, et opérations

## Pourquoi Hortonworks pour Hadoop ?

Fondée en 2011 par 24 ingénieurs venant de la société Yahoo! d'origine Constituée d'équipes de développement et d'exploitation Hadoop, Hortonworks est l'entreprise qui regroupe le maximum d'expérience sur Hadoop. Nos collaborateurs sont des participants et des leaders actifs dans le développement de Hadoop : ils conçoivent, bâtissent et testent le cœur de la plate-forme Hadoop. Nous avons cumulé des années d'expérience opérationnelle sur Hadoop et nous sommes les mieux armés pour supporter votre projet critique Hadoop.

Pour obtenir une analyse indépendante de la plate-forme de données Hortonworks et vous faire confirmer sa prééminence auprès des fournisseurs d'Apache Hadoop, vous pouvez télécharger le rapport *Forrester Wave™: Big Data Hadoop Solutions, Q1 2014* de Forrester Research.

## À propos de Hortonworks

Hortonworks développe, distribue et supporte la seule plate-forme de données Apache Hadoop 100 % open source. Notre équipe comprend les plus importants effectifs de concepteurs et d'architectes de l'écosystème Hadoop : au sein de cette communauté, ils représentent et pilotent les exigences de l'entreprise au sens large. La plate-forme de données Hortonworks s'intègre parfaitement avec les investissements informatiques existants ; les entreprises peuvent s'en servir pour bâtir et déployer des applications basées sur Hadoop. Hortonworks a des relations approfondies avec les principaux fournisseurs de centres de données stratégiques, ce qui permet à nos clients de tirer le meilleur profit de Hadoop. Pour plus d'informations, consultez [www.hortonworks.com](http://www.hortonworks.com).