



Points clés

- Répond à différents cas d'utilisation de données au repos
- Alimente des applications client en temps réel
- Fournit des analyses robustes

Plateforme de données Hortonworks

Une plateforme à architecture ouverte pour gérer les données en mouvement et au repos

Chaque entreprise est à présent reliée aux données mobiles Les données sont le futur et l'atout le plus précieux de votre organisation. La plateforme de données Hortonworks (HDP) est une distribution Apache Hadoop open source adaptée à l'entreprise et hautement sécurisée reposant sur une architecture centralisée (YARN). HDP répond aux besoins des données au repos, optimise les applications clients en temps réel et fournit des analyses robustes permettant d'accélérer la prise de décision et l'innovation.

La différence Hortonworks

HDP aide les entreprises à transformer leurs activités en débloquent le plein potentiel des big data avec les avantages suivants :

Ouvert	Central	Interopérable	Adapté à l'entreprise
HDP est composé de nombreux projets d'Apache Software Foundation (ASF) qui permettent aux entreprises de déployer, d'intégrer et de travailler avec des volumes sans précédent de données structurées et non-structurées. L'approche d'ASF consiste à fournir des logiciels de qualité professionnelle qui favorisent l'innovation et empêchent l'enfermement propriétaire.	YARN est le centre d'architecture d'Open-Enterprise Hadoop. Il alloue des ressources entre les diverses applications qui traitent des données. YARN coordonne les services à l'échelon du cluster pour les opérations, la gouvernance des données et la sécurité. YARN maximise également l'ingestion de données en permettant aux entreprises d'analyser les données de façon à prendre en charge divers cas d'utilisation. Ce processus permet aux opérateurs Hadoop d'étendre en toute confiance leurs ressources de big data à l'audience la plus large possible au sein de leurs organisations.	Son architecture open source à 100% permet à HDP d'être interopérable avec une large gamme d'applications de centre de données et de veille d'affaires. L'interopérabilité de HDP permet de minimiser les dépenses et les efforts nécessaires pour connecter les infrastructures informatiques des clients aux capacités de traitement et de données de HDP. Avec HDP, les clients peuvent conserver leur investissement dans l'architecture informatique existante lorsqu'ils adoptent Hadoop.	HDP est conçu pour les entreprises. Open-entreprise Hadoop fournit un fonctionnement cohérent, avec une gestion centralisée et une surveillance des clusters à travers un écran unique. Avec HDP, la sécurité et la gouvernance sont intégrées à la plateforme. Cette fonctionnalité permet de fournir un environnement sécurisé riche uniformément administré à travers les moteurs d'accès aux données.



La plateforme de données Hortonworks

HDP offre une distribution Hadoop open source, adaptée à l'entreprise et hautement sécurisée fondée sur une architecture centralisée. HDP répond à différents cas d'utilisation de données au repos, optimise les applications clients en temps réel et fournit des analyses robustes permettant d'accélérer la prise de décision et l'innovation.

Gestion des données

Les composants fondamentaux de HDP sont Apache Hadoop YARN et le système de fichier distribué d'Apache Hadoop (HDFS). Tandis que HDFS fournit un stockage évolutif, économique et tolérant les défaillances pour un grand lac de données, YARN propose l'architecture centralisée qui permet aux organisations de traiter plusieurs charges de travail simultanément. YARN fournit également la gestion

des ressources et l'architecture enfichable afin d'activer une grande variété de méthodes d'accès aux données.

Accès aux données

Avec YARN en son centre d'architecture, HDP fournit différents moteurs de traitement qui permettent aux utilisateurs d'interagir avec des données simultanément et de plusieurs façons. YARN permet la coexistence de différentes méthodes d'accès dans le même cluster contrairement aux ensembles de données partagées. Cette fonctionnalité évite les silos de données inutiles et coûteux. HDP permet l'utilisation de plusieurs moteurs de traitement de données, allant du langage de requête structuré interactif (SQL) et du streaming en temps réel à la science des données et au traitement par lots pour utiliser les données stockées sur une seule plateforme.

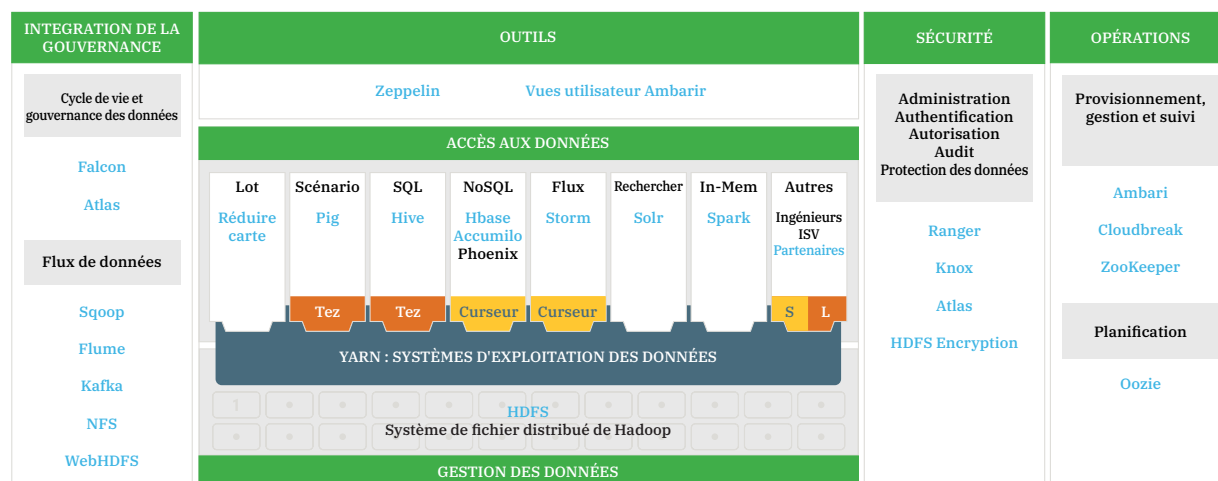


Figure 1 : La sécurité Hadoop de nouvelle génération

Sécurité et gouvernance

À mesure que les organisations mettent en application les initiatives Hadoop pour saisir de nouvelles opportunités en matière d'informations tirées de données, la gouvernance des données et les exigences de sécurité peuvent constituer un défi de taille. En réponse à ce défi, un consortium de leaders interprofessionnels nommé Data Governance Initiative (DGI) a été créé pour répondre à la nécessité d'une solution de gouvernance open source visant à gérer la classification, l'origine et la sécurité des données ainsi que la gestion du cycle de vie.

Apache Atlas, créé dans le cadre de DGI, permet aux organisations d'appliquer une classification cohérente des données dans l'écosystème de données. Apache Ranger fournit une administration centralisée de gestion de la sécurité pour Hadoop. En intégrant Atlas à Ranger, Hortonworks permet aux entreprises de mettre en place des politiques d'accès dynamiques lors de l'exécution qui permettent de prévenir les violations de manière proactive.

Cette intégration permet aux entreprises de mettre en œuvre des stratégies de sécurité dynamiques reposant sur la classification. La plateforme centralisée de Ranger permet aux administrateurs de données de définir une politique de sécurité basée sur des balises de métadonnées ou des attributs Atlas. Ils peuvent ensuite appliquer cette politique en temps réel à l'ensemble de la hiérarchie des ressources de données, y compris les bases de données, les tables et les colonnes.

Sécurité

Un lac de données alimenté par Hadoop peut constituer une base solide pour une nouvelle génération d'analyses et d'informations. Il est toutefois important de sécuriser les données avant de démarrer ou d'étendre une initiative Hadoop. En veillant à ce que la protection des

données et la gouvernance soient intégrées dans leurs grands environnements de big data, les entreprises peuvent tirer pleinement parti des analyses avancées sans exposer leur activité à de nouveaux risques.

Gouvernance

À mesure que les organisations mettent en application les initiatives Hadoop pour saisir de nouvelles opportunités en matière d'informations tirées de données, les exigences concernant la gouvernance des données peuvent constituer un défi de taille. La gestion de l'information visant à en identifier la valeur et permettre d'assurer le bon contrôle, la sécurité et la conformité des données de clients et d'entreprises est une exigence essentielle pour les architectures traditionnelles et les architectures de big data.

Fonctionnement

Le fonctionnement de HDP est conçu pour permettre aux services informatiques de mettre en ligne Hadoop rapidement en éliminant les approximations liées aux processus manuels et en les remplaçant par de meilleures pratiques automatisées et préconfigurées, des configurations guidées et un contrôle total des opérations. Le fonctionnement de HDP permet de simplifier l'exploitation des moteurs d'accès distribués multi-utilisateurs, multi-locataires et multidata et de gérer les clusters HDP à l'échelle via une interface utilisateur Web intégrée ou un écran unique.

HDP utilise Apache Ambari, une plateforme de gestion open source visant à provisionner, gérer, surveiller et sécuriser les clusters Hadoop. Ambari supprime les tâches manuelles souvent sujettes aux erreurs associées à l'utilisation de Hadoop. Il fournit également les points d'intégration nécessaires pour s'adapter de manière transparente à l'entreprise.

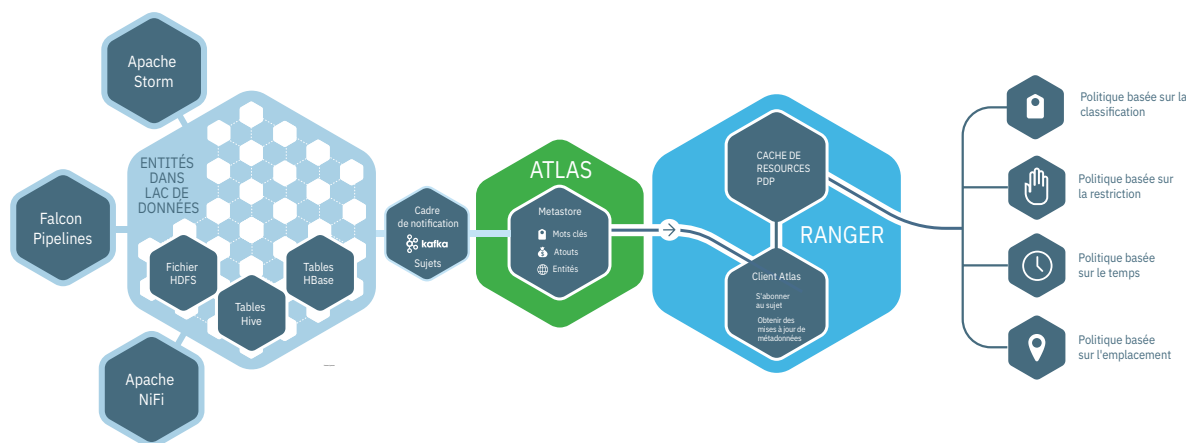


Figure 2 : La sécurité Hadoop de nouvelle génération

Options de déploiement

HDP offre différents choix d'infrastructure afin de déployer une plateforme de données ouverte et flexible. Les utilisateurs ont la possibilité de combiner les options d'infrastructure qui conviennent le mieux à leurs cas d'utilisation spécifiques.

Sur site

De nombreuses organisations ayant investi dans l'infrastructure de centre de données et les services administrés et qui envisagent à présent les fonctionnalités Hadoop considéreront l'implémentation sur site comme une option viable. HDP est conçu pour être facilement déployé sur site afin de s'intégrer aux centres de données existants.

Cloud

HDP peut être déployé dans le cloud dans le cadre d'Azure HDInsight de Microsoft. Azure HDInsight est une offre de services administrés sur le cloud de Microsoft Azure optimisée par HDP. Cette option de déploiement permet aux organisations de passer de téraoctets à pétaoctets (PB) en ce qui concerne les données à la demande en faisant tourner un nombre quelconque de nœuds à tout moment. Avec HDInsight, les entreprises peuvent également connecter leurs clusters Hadoop locaux au cloud.

Nuage hybride et Cloudbreak

Cloudbreak est une solution pour provisionner des clusters Hadoop sur une infrastructure en nuages. Cloudbreak, qui fait partie de HDP et qui est optimisé par Apache Ambari, aide les entreprises à simplifier l'approvisionnement des clusters dans le cloud et à optimiser l'utilisation des ressources cloud avec une mise à échelle flexible. Il est conçu pour les clients disposant d'un déploiement Hadoop sur site et souhaitant configurer des clusters dans le cloud plus facilement. Avec Cloudbreak, les clients peuvent choisir le fournisseur de cloud de leur choix et laisser Cloudbreak configurer le cluster dans le cloud.

HDP pour les équipes

Le déploiement réussi de Hadoop dans n'importe quelle organisation dépend de l'utilisation d'ensembles de compétences et de ressources existantes pour adopter l'architecture des big data. HDP fournit des outils précieux et attribue des compétences à tous les niveaux de votre équipe en charge des big data.

L'expert en données massives

Apache Spark, qui fait partie de HDP, joue un rôle important en matière de science des données. Les experts en données massives utilisent couramment l'apprentissage machine, un ensemble de techniques et d'algorithmes ayant la capacité de tirer des enseignements des données. Ces algorithmes sont souvent itératifs et la capacité de Spark à mettre en cache les données en mémoire accélère considérablement le traitement itératif des données, ce qui en fait un moteur de traitement idéal pour la mise en œuvre d'algorithmes de ce genre.

L'analyste commercial

HDP fournit aux analystes commerciaux un accès rapide à de grandes quantités de données via SQL sur les interfaces Hadoop fournies par Apache Hive, Spark SQL et Apache Phoenix. Grâce à ces interfaces, les analystes commerciaux peuvent utiliser leurs outils de renseignements d'affaires et d'analyse décisionnelle préférés pour créer des rapports, des visualisations, des tableaux de bord et des feuilles de résultats afin de prendre des décisions éclairées et plus efficaces.

Le développeur

HDP fournit un ensemble d'outils variés pour les développeurs d'applications, tels que Spark et Apache Zeppelin. HDP propose également un ensemble d'interfaces de programme d'application (API) natives qui facilitent le développement : WebHDFS fournit une interface REST pour écrire, manipuler et supprimer les fichiers HDFS, tandis que WebHCAT fournit un point d'intégration essentiel pour accéder aux métadonnées et au schéma des données Hadoop.

Opérateurs Hadoop

Grâce à Ambari, les opérateurs Hadoop bénéficient des avantages suivants pour rationaliser les opérations Hadoop :

- Mises à niveau plus flexibles : Ambari offre un moyen plus rapide de mettre à niveau les clusters en automatisant la maintenance et les nouvelles fonctionnalités lorsque le cluster est en panne
- Opérations de sécurité simplifiées : Les configurations de service pour Ranger fournissent une continuation de la nouvelle expérience utilisateur. Le stockage facultatif des informations d'identification Kerberos et les paramètres de sécurité personnalisables simplifient l'administration et offrent un environnement sécurisé.
- Dépannage amélioré : Ambari fournit un affichage graphique widget métrique personnalisable ainsi que la possibilité d'exporter des indicateurs pour identifier et répondre aux problèmes rapidement.

Secteurs

HDP permet aux entreprises de divers secteurs d'améliorer leurs principales fonctions :

Services financiers	Télécommunications
<ul style="list-style-type: none"> • Gérer le risque de défaut • Améliorer les ventes croisées aux clients • Détecter le blanchiment d'argent. 	<ul style="list-style-type: none"> • Analyser les enregistrements des détails des appels (CDR) • Servir proactivement les infrastructures de service • Rationaliser les investissements en infrastructure • Développer de nouveaux produits et services.
Vente au détail	Pétrole et gaz
<ul style="list-style-type: none"> • Construire une vue à 360° de leurs clients • Localiser et personnaliser les expériences de consommation • Gérer les chaînes d'approvisionnement efficacement • Comprendre les changements de sentiment envers la marque à travers l'analyse des sentiments • Optimiser les sites Web, les campagnes et l'agencement des magasins. 	<ul style="list-style-type: none"> • Surveiller la production en amont dans des régions éloignées • Ralentir les courbes de déclin • Réparer de manière proactive l'équipement de valeur • Établir des rapports sur la conformité avec les règlements en matière de santé et de sécurité environnementale.

Conclusion

Les données constituent un outil fondamental pour toutes les entreprises, tous secteurs confondus. Elles déterminent la façon dont vous développez de nouveaux produits et services, l'efficacité avec laquelle vous opérez, etc. HDP gère les données en mouvement en aidant à acquérir et à transporter des données en toute sécurité vers le HDP. HDP gère les données au repos pour pratiquement tous les types de données, avec une gouvernance, une sécurité et un fonctionnement de niveau professionnel, ce qui permet ainsi à votre entreprise de rester compétitive.

Pourquoi IBM ?

IBM® est une société multinationale américaine de conseils et d'informatique dont le siège est à Armonk, New York. IBM fabrique et commercialise du matériel informatique, des logiciels et des services d'hébergement et de consultation dans des domaines allant des ordinateurs centraux aux nanotechnologies. IBM est également un important organisme de recherche, détenant en 2017 le record de la plupart des brevets générés par une entreprise pendant 24 années consécutives.

Pour en savoir plus

Pour en savoir plus sur la plateforme de données Hortonworks, contactez votre représentant IBM ou votre partenaire commercial IBM, ou consultez le site à l'adresse suivante : ibm.com/hadoop



IBM France
17 Avenue de l'Europe
92275 Bois Colombes Cedex

IBM, le logo IBM, **ibm.com** sont des marques ou des marques déposées d'International Business Machines Corp. dans de nombreux pays du monde. Les autres noms de produits et de services peuvent être des marques d'IBM ou d'autres sociétés. Une liste actualisée des marques déposées IBM est accessible sur le web sous la mention « Copyright and trademark information » à l'adresse ibm.com/legal/copytrade.shtml

Microsoft et le logo de Windows sont des marques de Microsoft Corporation aux États-Unis et/ou dans d'autres pays.

Les exemples de clients fournis ne sont mentionnés qu'à titre d'information. Les performances réelles peuvent varier selon les configurations et les conditions de fonctionnement spécifiques. Il appartient à l'utilisateur d'évaluer et de vérifier le fonctionnement d'autres produits ou programmes avec les produits et programmes d'IBM. LES INFORMATIONS CONTENUES DANS CE DOCUMENT SONT FOURNIES « EN L'ÉTAT », SANS AUCUNE GARANTIE EXPRESSE OU TACITE, NOTAMMENT SANS AUCUNE GARANTIE DE QUALITÉ MARCHANDE OU D'ADAPTATION À UN EMPLOI SPÉCIFIQUE, ET SANS AUCUNE GARANTIE OU CONDITION DE NON INFRACTION VIS-À-VIS DES LOIS. Les produits IBM bénéficient d'une garantie, conformément aux conditions générales des contrats dans le cadre desquels ils sont fournis.

© Copyright IBM Corporation 2018



Veillez recycler