

# Maîtriser les informations dans votre écosystème de données évolutif



*Les nouveautés de la plateforme InfoSphere Information Server V11.5*

---

## Points clés

- Permet d'obtenir de meilleurs résultats dans vos initiatives d'analyse et projets stratégiques.
  - Garantit que les informations des initiatives métier et stratégiques sont fiables, cohérentes et gérées en temps réel.
  - Permet de mieux comprendre, nettoyer, surveiller et transformer et diffuser les données.
- 

## Introduction

Les plateformes de données évolutives, comme Apache Hadoop, apportent des économies et des possibilités d'analyse inégalées. IBM® vous permet de tirer pleinement parti des fonctionnalités et des promesses de Hadoop pour obtenir de meilleurs résultats dans vos initiatives d'analyse et projets stratégiques. Avec les capacités de gestion de bout en bout des informations qu'apporte IBM InfoSphere Information Server V11.5, vous pouvez mieux comprendre, nettoyer, surveiller, transformer et distribuer les données. IBM contribue en outre à rapprocher les activités métier et l'informatique, grâce à une collaboration plus performante. Les fonctionnalités d'« intégration souple » d'InfoSphere Information Server assurent la sécurité, la cohérence et la gouvernance en temps réel des informations des initiatives métier et stratégiques : Big Data, analyse du point d'impact, gestion des données de référence et entrepôt de données.

D'emblée, InfoSphere Information Server a été conçu comme une plateforme de traitement massivement parallèle (MPP) capable de prendre en charge tous les volumes de données, quelles que soient leur taille et leur complexité. À ce jour sans équivalent, cet outil vous apporte la flexibilité – avec les processus ETL (Extract, Transform, Load) ou ELT (Extract, Load, Transform) –, les performances et l'évolutivité nécessaires à la réussite de vos projets.



InfoSphere Information Server V11.5 est disponible en quatre versions conçues pour aider votre entreprise à résoudre des problématiques clés en matière de gestion de l'information (figure 1) :

- **InfoSphere Information Server for Data Integration :** transformer en souplesse toutes sortes de données et les distribuer à tous les systèmes possibles, en accélérant la création de valeur et en réduisant le risque informatique.
- **InfoSphere Information Server for Data Quality :** établir et gérer des données de qualité, en transformant le déluge de données en informations fiables.
- **InfoSphere Information Governance Catalog :** mieux comprendre les données et favoriser la collaboration entre l'équipe informatique et les équipes produits pour améliorer la communication et créer un cadre de gouvernance de l'information.

- **InfoSphere Information Server Enterprise Edition :** profiter des fonctionnalités de trois versions dans une offre complète, pour engager une démarche d'intégration des informations dans un domaine avant de l'étendre en fonction des besoins à d'autres domaines en vue d'optimiser les résultats.

Votre entreprise peut choisir la version qui correspond le mieux à ses besoins ou opter pour la version complète et profiter de toutes les fonctionnalités (figure 2).

### InfoSphere Information Server for Data Integration

Les fonctionnalités d'InfoSphere Information Server vous permettent non seulement de comprendre, nettoyer, surveiller, transformer et diffuser vos données, mais aussi de collaborer pour rapprocher les activités métier et l'informatique.

InfoSphere Information Server for Data Integration :	InfoSphere Information Server for Data Quality :	InfoSphere Information Governance Catalog :	InfoSphere Information Server Enterprise Edition :
<p>Intégration des données dans Hadoop</p> <p>Connectivité des fichiers HDFS</p> <p>Traitement par décomposition descendante (Push-Down) dans les bases de données relationnelles</p> <p>Protection intégrée des données sensibles en transit grâce à InfoSphere Optim Masking.</p>	<p>Amélioration des performances de classification des données</p> <p>InfoSphere Information Analyzer</p> <p>Gestion de la qualité des données par le nettoyage et l'enrichissement</p>	<p>Architecture évolutive de Governance Catalog</p> <p>Définition de classes de données</p> <p>Fonctionnement en plusieurs langues</p> <p>Prise en charge des définitions de schéma XML</p> <p>Possibilité d'échange d'actifs pour des contenus étendus de lignage</p>	<p>Intégration des données dans Hadoop</p> <p>Connectivité des fichiers HDFS</p> <p>Traitement par décomposition descendante (Push-Down) dans les bases de données relationnelles</p> <p>Protection intégrée des données sensibles en transit grâce à</p> <p>InfoSphere Optim Masking Amélioration des performances de classification des données</p> <p>InfoSphere Information Analyzer</p> <p>Gestion de la qualité des données par le nettoyage et l'enrichissement</p> <p>Architecture évolutive de Governance Catalog</p> <p>Définition de classes de données</p> <p>Fonctionnement en plusieurs langues</p> <p>Prise en charge des définitions de schéma XML</p> <p>Possibilité d'échange d'actifs pour des contenus étendus de lignage</p>

Figure 1. Les quatre versions d'IBM InfoSphere Information Server V11.5 qui répondent à un large éventail de problématiques.

Ces nouvelles fonctionnalités incluent entre autres :

- **Une «évolutivité accrue et de hautes performances, permettant d'accéder plus rapidement aux données nettoyées** : un moteur de traitement massivement parallèle (MPP) qui s'exécute en mode natif dans Hadoop – où les données résident – propose un éventail complet de fonctionnalités d'intégration et de gouvernance. Il aide votre organisation à atteindre des hauts niveaux de performance et de vitesse pour les activités d'intégration et de gouvernance des données : connectivité, transformation, nettoyage, enrichissement et diffusion.
- **Toutes les fonctionnalités de connectivité, de transformation et de diffusion des données s'exécutent désormais en mode natif dans Hadoop** : vous bénéficiez également d'un accès étendu et simplifié aux fichiers HDFS (Hadoop Distributed File System) dans différents formats et jeux de caractères, ainsi que de fonctionnalités de sécurité stratégiques telles que Kerberos et les passerelles sécurisées.
- **Faible coût d'intégration grâce à une productivité accrue** : avec le moteur d'InfoSphere Information Server, vos développeurs peuvent utiliser les fonctionnalités de l'outil de conception graphique intégré pour réduire considérablement le coût d'intégration par rapport à un codage manuel.
- **Fonctionnalités supplémentaires sans surcoût** : votre organisation bénéficie de fonctionnalités supplémentaires, comme l'accès à des applications professionnelles prises en charge, comme :
  - Salesforce.com
  - Hyperion Essbase
  - Siebel
  - JD Edwards Enterprise One
  - PeopleSoft Enterprise
  - Applications Oracle®

Vous pouvez également tirer parti de l'optimisation du traitement par décomposition descendante grâce aux fonctionnalités d'IBM DataStage Balanced Optimisation. Ainsi, vous exploitez au maximum les capacités et la puissance de calcul disponibles dans vos bases de données relationnelles et dans InfoSphere DataStage.

## InfoSphere Information Server for Data Quality

La qualité des données est un aspect essentiel de la gouvernance des informations, qui est au cœur de la plateforme InfoSphere Information Server, car elle permet d'obtenir des informations cohérentes, précises et sécurisées. InfoSphere Information Server propose un large éventail de fonctionnalités : profilage, standardisation, enrichissement, suivi actif de la qualité des données.

### Découvrir où les données sensibles sont stockées

Les organisations doivent être responsables et diligentes en matière de protection des données personnelles. Le point de départ consiste à déterminer où ces données sont stockées. La solution InfoSphere Information Server for Data Quality assure la confidentialité, le masquage et la gestion des données de test en identifiant l'emplacement des données sensibles, des informations personnelles et d'autres classes de données.

- Accélérez la création de valeur en identifiant le type des données dans un tableau ou un fichier à l'aide de plusieurs dizaines de classes de données prédéfinies : numéro de carte de crédit, numéro fiscal, numéro de téléphone, adresse électronique, etc.
- Ajoutez d'autres classes propres à votre organisation. Créez et personnalisez trois types de classes de données évolutives : liste de valeurs valides, expression rationnelle (Regex) et Java.
- Donnez à votre analyste un contrôle total. Dans ses résultats, l'analyse des colonnes propose des classifications en fonction des valeurs des données. Les utilisateurs peuvent examiner ces classifications et, si nécessaire, choisir une autre classe plus représentative des données stockées dans chaque colonne.

### Performances accrues pour répondre aux demandes croissantes concernant les données

Tirez parti d'une évolutivité accrue et d'une consommation de ressources moindre pendant le profilage et l'analyse des données. InfoSphere Information Server for Data Quality propose une nouvelle fonctionnalité de limitation du nombre de valeurs distinctes stockées, ce qui permet de réduire considérablement la taille de la base de données des résultats et d'accélérer le profilage des données. Cette plateforme est également conçue pour favoriser l'évolutivité de l'analyse des clés étrangères et interdomaines. Ses nouvelles méthodes ne nécessitent plus l'analyse complète d'une colonne avec capture de toutes les valeurs distinctes.

### Le choix de l'emplacement d'exécution

Toutes les fonctionnalités permettant d'évaluer et de surveiller la qualité de vos données s'exécutent désormais directement dans un cluster Hadoop. Vous pouvez exécuter le même profilage et les mêmes règles de données dans ou hors de votre environnement Hadoop, notamment :

- Analyse des données
- Standardisation
- Correspondance
- Survie
- IBM InfoSphere QualityStage Module for US Address Certification
- IBM InfoSphere QualityStage Module for Address Verification Interface

### InfoSphere Information Governance Catalog

La mise en conformité automatisée des données vous permet de trouver, comprendre, sécuriser et d'exploiter les meilleures données disponibles, plus rapidement qu'auparavant. Les fonctionnalités de lignage d'InfoSphere Information Governance Catalog collectent automatiquement les sources et les mouvements des données. Donnez à tous vos utilisateurs des outils professionnels permettant de comprendre en langage naturel les informations, ainsi que des indicateurs de qualité pour qu'ils puissent identifier les données fiables. La détection en temps réel des données privées et sensibles facilite la mise en conformité rapide vis-à-vis des réglementations et des exigences de l'entreprise.

#### Proposez une vue complète des sources et mouvements des données au sein de votre entreprise, y compris des sources de données héritées et propriétaires

- Définissez les propriétés d'affichage, de structure et de confinement permettant d'enregistrer de nouveaux actifs informationnels dans le catalogue.
- Automatisez l'importation et l'ajout dans le catalogue de nouveaux actifs informationnels via l'interface de programmation d'application (API) IGC REST.

- Recherchez et affichez les nouveaux actifs informationnels, et assurez-en la gouvernance.
- Automatisez la mise en conformité des données privées ou sensibles en scrutant et en détectant automatiquement les classes de données et les statistiques de profilage.
- Affichez les classifications de données détectées, publiées par IBM InfoSphere Information Analyser, et sélectionnez la classe de données standard ou attendue.
- Identifiez les éléments contenant des informations sensibles ou à accès restreint, en évaluant leur classe de données.
- Assurez la gouvernance des classes de données en les associant aux règles qui doivent contrôler ou restreindre l'accès à l'élément de données.
- Créez et gérez les classes de données, comme l'expression rationnelle, la liste de valeurs valides ou la plage.

#### Déployez le catalogue auprès de tous les utilisateurs de votre entreprise avec une interface multilingue et localisée

- Les utilisateurs peuvent choisir la langue d'affichage et de recherche de définitions métier dans le catalogue.
- Ils peuvent consulter la langue définie pour les définitions métier, les définitions métier associées et les actifs partagés affectés.
- Les administrateurs peuvent gérer et définir l'ensemble des langues prises en charge pour les définitions métier.

#### Permettez à votre organisation d'inclure vos sources de données les plus importantes, comme les modèles XSD, dans votre modèle de gouvernance

- Affichez et explorez les définitions de schéma XML (modèles XSD) et assurez-en la gouvernance.
- Les administrateurs peuvent importer des modèles XSD grâce à IBM InfoSphere Metadata Asset Manager et capturer les éléments, attributs et types XSD.
- Les utilisateurs peuvent rechercher, afficher et modifier les détails d'un modèle XSD.

## InfoSphere Information Server Enterprise Edition

Cette offre conjugue les fonctionnalités des versions InfoSphere Information Server for Data Integration, InfoSphere Information Server for Data Quality et InfoSphere Information Governance Catalog pour gagner en flexibilité et en évolutivité.

### Protection intégrée des données sensibles en transit grâce à InfoSphere Optim Masking

La licence InfoSphere Information Server Enterprise Edition intègre aujourd'hui les bibliothèques d'InfoSphere Optim Data Masking, qui protègent les informations sensibles lors d'une tâche d'intégration ou de nettoyage de données exécutée en parallèle.

### Exécution en mode natif dans Hadoop : une évolutivité sans limites

InfoSphere Information Server propose des fonctionnalités révolutionnaires qui ouvrent la voie à des modes d'utilisation de Hadoop inédits. Dans cette version, InfoSphere Information Server peut s'exécuter directement dans un cluster Hadoop. Cela signifie que toutes les fonctionnalités de connectivité, de transformation, de nettoyage, d'amélioration et de distribution des données, utilisées par des milliers d'entreprises depuis des années, sont immédiatement disponibles sur la plateforme Hadoop.

Vous bénéficiez ainsi d'une palette complète de fonctionnalités d'intégration et de gouvernance, permettant de relever certains des défis les plus complexes de l'industrie et conçues pour s'exécuter là où votre organisation stocke ses données. Ces fonctionnalités minimisent les mouvements de données, accélérant la création de valeur métier.

- Adaptez vos charges de travail à tous les nœuds de données de vos environnements Hadoop.
- Distribuez automatiquement les bibliothèques d'exécution à l'ensemble des nœuds de données Hadoop.
- Exploitez des fonctionnalités telles que les intitulés de nœud Hadoop et les files d'attente de planification Yarn pour localiser les ressources.
- Utilisez la localité des données HDFS pour en optimiser.
- Tirez parti automatiquement des fonctionnalités en mémoire, comme la gestion des pipelines, le partitionnement et le repartitionnement dynamique des données, pour minimiser le temps de traitement des charges de travail.
- Bénéficiez de la prise en charge des métadonnées de conception et d'exécution, ainsi que du lignage intégré des données dans vos charges de travail Hadoop.
- Utilisez les clusters compatibles Kerberos.

### Niveau de version Hadoop

Comme dans les versions précédentes, InfoSphere Information Server V11.5 reconnaît les distributions Hadoop conformes aux spécifications Open Data Platform (ODP), notamment IBM BigInsights et HortonWorks Data Platform. Cette plateforme reconnaît aussi Cloudera Distribution for Hadoop.

IBM prend en charge ces distributions Hadoop pour la lecture et l'écriture de données, ainsi que le déploiement du moteur InfoSphere Information Server pour exécuter des opérations de gouvernance et d'intégration des données en mode natif dans Hadoop.

### Mise à jour du référentiel et des options du serveur d'applications

Le niveau Référentiel est maintenant compatible avec Oracle 12g (dont RAC) et SQL Server 2014. Quant au niveau Services, il est compatible avec IBM WebSphere® Application Server 8.5.5. Pour en savoir plus sur la configuration système requise, visitez : [ibm.com/support/docview.wss?uid=swg27046843](http://ibm.com/support/docview.wss?uid=swg27046843)

### Mise à jour des fonctionnalités de connectivité

InfoSphere Information Server V11.5 étend sa connectivité, déjà riche, aux sources de données d'entreprise. Son nouveau connecteur gère la connectivité et les métadonnées des fichiers HDFS. De plus, cette plateforme prend désormais en charge d'autres types de fichiers, d'autres types de données, le lignage automatique des données au niveau des champs et l'importation des métadonnées des fichiers.

Les autres mises à jour (connecteurs et phases) compatibles avec les dernières versions de ces sources de données sont les suivantes :

- Phase SQL Server pour Microsoft® SQL Server 2014
- Phase Sybase pour Sybase ASE et Sybase IQ, Version 16
- Teradata Connector pour Teradata Database 15.10
- IBM Netezza Connector pour Netezza Performance Server 7.1
- Connecteur IBM WebSphere ILOG JRules pour WODM 8.7



## Pourquoi IBM ?

L'offre IBM Information Integration and Governance (IIG) apporte des fonctionnalités de pointe nécessaires pour faire face aux défis du Big Data. Elle se caractérise par son évolutivité et ses performances optimales sur les volumes massifs de données, par son agilité, son niveau idéal d'intégration et sa gouvernance qui permettent d'accélérer la mise à disposition des données, mais aussi par la prise en charge et la protection d'un large éventail de types de données et de systèmes Big Data. IBM InfoSphere contribue à la réussite des projets d'analyse et de Big Data en apportant aux utilisateurs la confiance nécessaire pour transformer les données en informations.

## Pour en savoir plus

Pour en savoir plus sur IBM InfoSphere Information Server V11.5, contactez votre représentant ou votre partenaire commercial IBM, ou visitez le site :

[ibm.com/software/products/en/infosphere-information-server](http://ibm.com/software/products/en/infosphere-information-server)

- De plus, IBM Global Financing propose de multiples solutions pour vous aider à acquérir la technologie dont vous avez besoin pour développer votre activité. Nous assurons la gestion du cycle de vie complet des produits et services informatiques, depuis l'acquisition jusqu'à la mise au rebut.

Pour en savoir plus, visitez le site :

[ibm.com/financing](http://ibm.com/financing)

### IBM France

17 Avenue de l'Europe  
92275 Bois Colombes Cedex

IBM, le logo IBM, ibm.com, AIX, BigInsights, DataStage, DB2, ILOG, InfoSphere, Optim, QualityStage, System z et WebSphere sont des marques d'International Business Machines Corp., déposées 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](http://ibm.com/legal/copytrade.shtml).

Java et les marques et logos contenant Java sont des marques ou des marques déposées d'Oracle et/ou de ses filiales.

Linux est une marque déposée de Linus Torvalds aux États-Unis et/ou dans d'autres pays.

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

Ce document est considéré comme à jour à sa date initiale de publication et peut être modifié par IBM à tout moment. Toutes les offres ne sont pas disponibles dans chaque pays dans lequel opère 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 2016



Veillez recycler