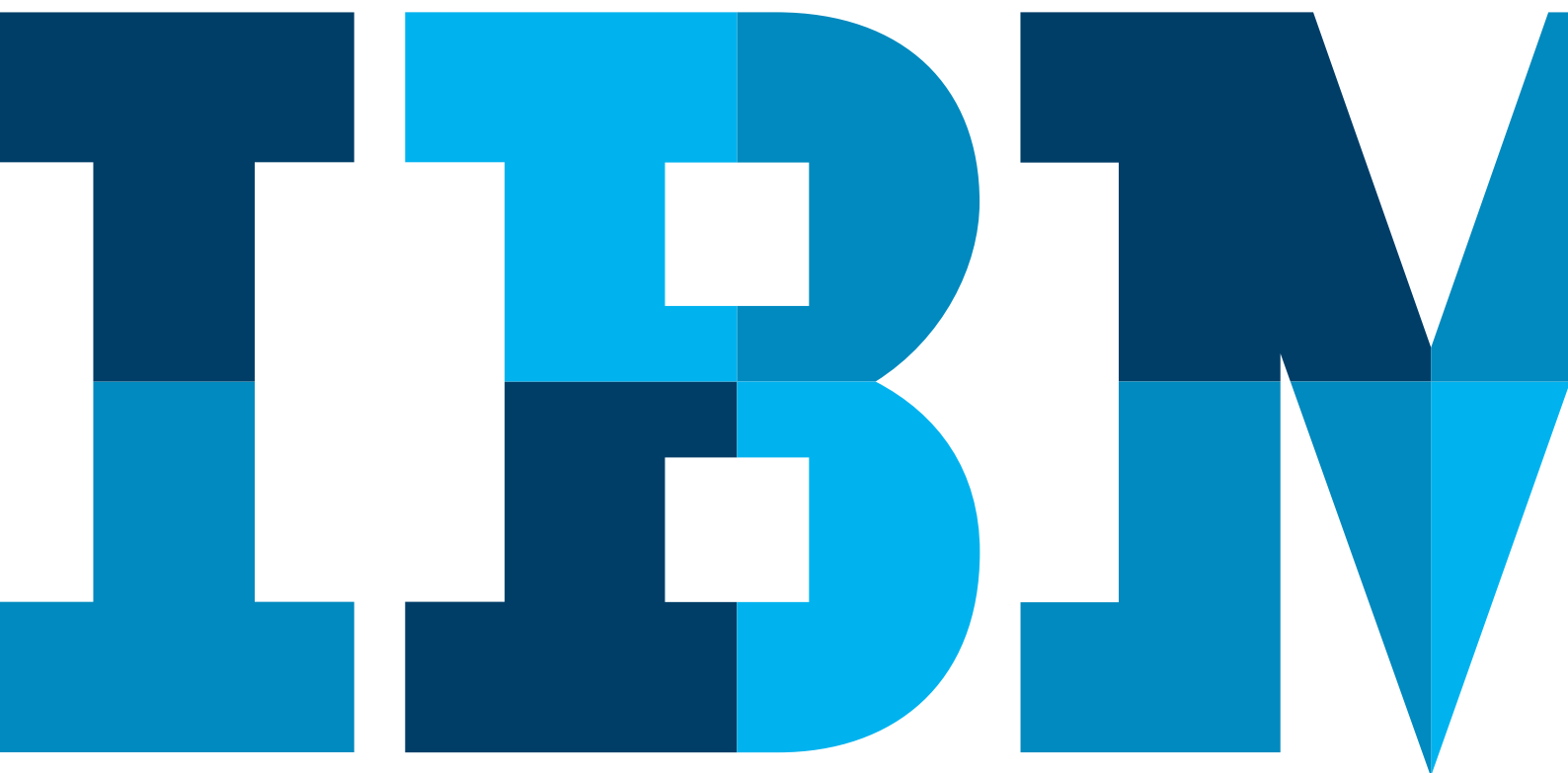


# Maîtriser la puissance des outils d'analyse avancée avec IBM PureData System for Analytics

*Une approche fondée sur une appliance pour simplifier l'accès aux analyses avancées*



## Sommaire

- 2 Introduction
- 3 Qu'est-ce que la solution IBM® Netezza Analytics peut m'apporter ?
- 4 Comment utiliser la solution IBM Netezza Analytics ?
  - 4 Préparation et transformation des données
  - 4 Création et publication d'un modèle
  - 4 Évaluation par modèle
  - 4 Développement d'algorithmes
- 6 Récapitulatif
- 6 Découvrez par vous-même les capacités de la solution avec notre offre d'essai gratuite
- 6 La solution IBM PureData System for Analytics en bref
- 7 Les solutions IBM Data Warehousing and Analytics en bref
- 7 Pour de plus amples informations.

L'offre IBM PureData System for Analytics, associée à la solution IBM Netezza Analytics, permet de repousser les limites des outils analytiques en conjuguant notre appliance d'entrepôt de données exceptionnelle et une solution d'analyse apportant hautes performances et évolutivité. Les appliances PureData System for Analytics sont toutes équipées de la technologie Netezza, qui permet de traiter des volumes massifs de données pour résoudre des problèmes complexes, et ce, de manière considérablement plus rapide que les solutions habituelles. Intégrée avec chaque appliance PureData System for Analytics, la solution IBM Netezza Analytics constitue une plateforme d'analyse avancée, ouverte et flexible, dédiée au développement et au déploiement de traitements d'analyse en base de données. Vous disposez ainsi des moyens de produire des résultats pouvant changer la donne.

En intégrant les outils d'analyse dans ses appliances, IBM a simplifié le déploiement des solutions d'analyse tout en capitalisant sur la puissance de traitement de l'architecture associée.

Avec la solution IBM Netezza Analytics, les traitements d'analyse parallèle (préparation des données, data mining, modélisation prédictive, optimisation) bénéficient de l'architecture AMPP (Asymmetric Massively Parallel Processing) de l'appliance PureData System for Analytics pour assurer des analyses évoluées à haut débit, appliquées à des volumes massifs de données. Les appliances IBM PureData System for Analytics sont facilement extensibles pour créer des analyses puissantes, avancées et personnalisées, intégrables ensuite dans un dispositif plus global. Les applications analytiques et les outils de visualisation et de business intelligence permettent d'exploiter des fonctions d'analyse évoluées, basées sur des traitements parallèles, grâce à différentes méthodes de programmation, notamment les langages SQL, Java™, MapReduce, Perl, Python, Lua, R, C, C++ et Fortran, pour des analyses à la fois puissantes et pertinentes. L'environnement complet de la solution se caractérise par sa flexibilité, et vous permet d'utiliser vos outils favoris pour mettre en œuvre des analyses ad hoc, créer des prototypes et déployer vos analyses en production.

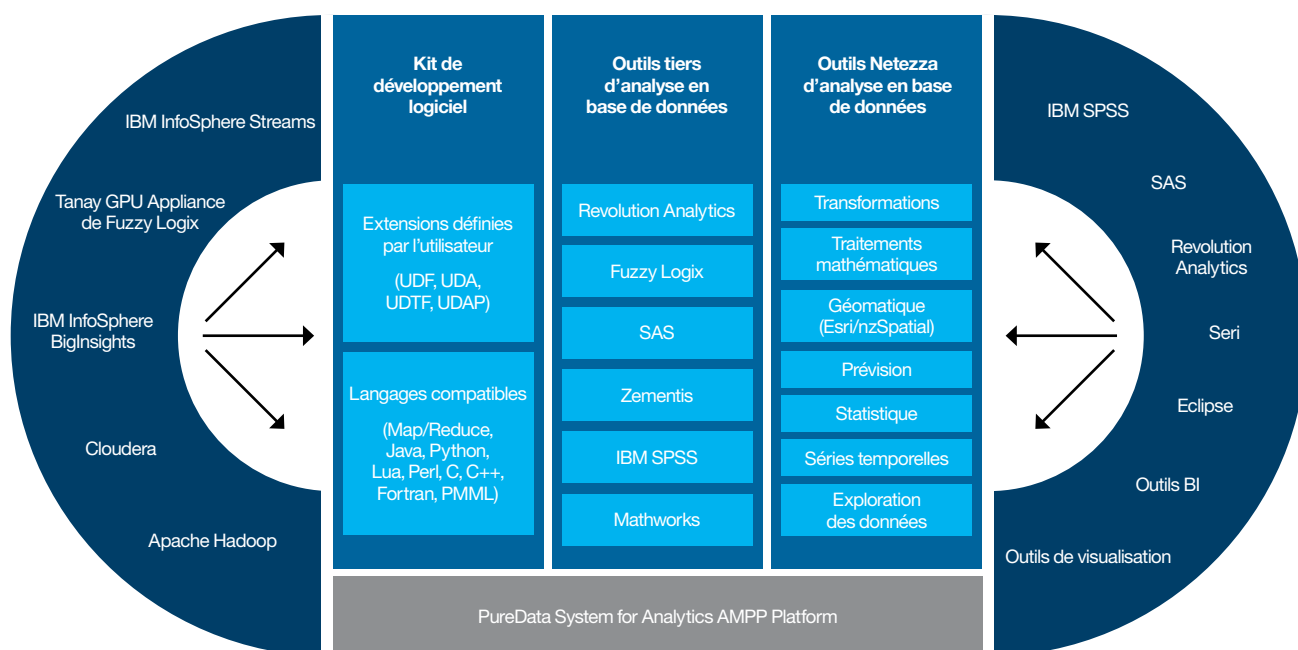


Figure 1. IBM Netezza Analytics 2.5

## Qu'est-ce que la solution IBM® Netezza Analytics peut m'apporter ?

Pour assurer les hautes performances et l'évolutivité des applications d'analyse, IBM Netezza Analytics exploite la convergence des traitements parallèles et des volumes massifs de données. L'architecture AMPP de l'appliance PureData System for Analytics permet des traitements analytiques capitalisant sur le volume des données et le parallélisme des tâches pour obtenir des résultats pertinents à une vitesse exceptionnelle. Grâce à l'intégration parfaite des analyses et d'une architecture AMPP inégalée, IBM a repoussé les limites de la technologie pour obtenir des performances accélérées et la capacité à effectuer des traitements de données impossibles jusqu'ici. Grâce à l'évolutivité de la solution, englobant aussi bien les volumes de données que la complexité des traitements, les utilisateurs (statisticiens, spécialistes des modèles, développeurs) disposent du bon outil, au bon moment, et peuvent déceler des opportunités à forte valeur ajoutée.

Il est de pratique courante, aujourd'hui, de copier les données à partir d'un entrepôt de données pour effectuer des analyses, ce qui conduit à de longs cycles de développement et à des coûts plus élevés de maintenance des logiciels. Ces opérations s'appuient sur des programmes destinés à extraire, transformer et charger les données par des échanges entre l'entrepôt de données et d'autres plateformes. L'appliance PureData System for Analytics, associée à la plateforme IBM Netezza Analytics, simplifie ce traitement en rapprochant les activités d'analyse des données stockées dans l'entrepôt, ce qui contribue à éliminer ou à réduire de manière significative ces étapes, dépourvues de valeur ajoutée. Grâce à cette simplification de l'environnement, les analystes peuvent se recentrer sur des activités à plus forte valeur ajoutée (modélisation, simulation, optimisation) plus rentables pour l'entreprise.

La solution IBM Netezza Analytics intègre un ensemble étendu de fonctions d'analyse évoluées, proposées par IBM et ses partenaires. Les développeurs d'applications et les spécialistes de la modélisation disposent ainsi de la bonne combinaison de fonctions d'analyse pour concevoir et développer une solution complète adaptée à leurs problématiques spécifiques. Les entreprises peuvent ainsi s'appuyer sur des outils d'analyse parallélisés et hautement performants, constituant l'assise de leur compétitivité, et leur permettant de développer plus rapidement des applications analytiques.

À titre d'exemple, dans le cadre d'une application de gestion des relations client (CRM), caractérisée par un volume énorme de transactions et une variabilité élevée de clients, la segmentation est fréquemment utilisée pour identifier les clients rentables, la modélisation prédictive servant à déterminer comment transformer en clients hautement profitables ceux qui ne le sont pas. La solution conjugue la modélisation et l'optimisation pour déterminer la communication optimale (conception et diffusion du message, promotion) permettant d'améliorer les taux de réponse. En outre, le data mining et la modélisation prédictive permettent d'identifier et d'accompagner les clients les plus fidèles, mais aussi de fidéliser davantage les autres en leur proposant des offres motivantes appropriées.

## Comment utiliser la solution IBM Netezza Analytics ?

La solution IBM Netezza Analytics est une plateforme intégrée d'analyse avancée, spécifiquement conçue pour des traitements analytiques efficaces. La plateforme IBM Netezza Analytics est utilisée pour :

- Construire et déployer des applications d'analyse évoluées ;
- Exploiter les outils d'analyse associés aux traitements parallèles avec des outils de visualisation ou de décisionnel (BI) ;
- Procéder à des analyses ad hoc, en particulier concernant les volumes massifs de données ou les problèmes nécessitant des traitements informatiques intensifs.

Les outils de visualisation et de business intelligence s'appuient sur les fonctionnalités d'analyse de la plateforme IBM Netezza Analytics, en y associant des processus SQL pour effectuer des recherches ciblées. Cependant, les développeurs d'application et les spécialistes de la modélisation demandent davantage de flexibilité pour créer des applicatifs ou des analyses ad hoc, et en particulier, les fonctionnalités suivantes :

- Préparation et transformation des données,
- Création et publication de modèles,
- Évaluation par modèle,
- Développement d'algorithmes.

### Préparation et transformation des données,

Un analyste, chargé des activités de préparation et de transformation des données, peut utiliser les outils et les utilitaires IBM Netezza Analytics pour faciliter les opérations. Les outils proposés sont les suivants :

- Statistiques descriptives,
- Nettoyage des données,
- Transformation des données,
- Sélection de caractéristiques et réduction dimensionnelle,
- Tests et utilitaires statistiques.

Lorsque des opérations de préparation et de transformation des données sont nécessaires, IBM propose des packages hautement performants pour accélérer le processus et étendre jusqu'à plusieurs pétaoctets les capacités en volume de données client stockables sur l'appliance.

## Création et publication d'un modèle d'analyse

La création d'un modèle d'analyse s'appuie généralement sur un processus itératif. Les données utilisées pour construire le modèle peuvent être préparées et transformées par un spécialiste des modèles (responsable de l'extraction des données, statisticien ou analyste quantitatif). Il est également possible que ce spécialiste commence par procéder à une analyse ad hoc, à la demande d'un responsable de l'entreprise. Une fois les résultats analysés, cet utilisateur peut créer une représentation affinée, proche d'un prototype, et permettant, au final, un déploiement en production. Le spécialiste des modèles développe souvent plusieurs modèles au cours de ce processus, puis évalue et compare leurs performances avant d'en choisir un pour le déployer. À chaque phase du cycle de développement par itérations, il peut utiliser différents outils de développement et d'analyse en fonction de l'évolution des besoins et des objectifs de performances.

Pour chacune des phases de ce processus, la plateforme IBM Netezza Analytics apporte ses caractéristiques de flexibilité. Le spécialiste des modèles peut créer ses modèles grâce à l'un des environnements de développement proposés ou utiliser directement les extensions suivantes :

- Plug-in IBM pour l'environnement Eclipse,
- Environnement intégré de développement (IDE) proposé par un partenaire commercial,
- Environnement de développement d'analyse (ADE) proposé par un partenaire commercial, permettant la création d'un workflow analytique,
- Langages de programmation compatibles avec la plateforme IBM Netezza – SQL, Java, MapReduce, Perl, Python, Lua, R, Fortran, C ou C++.

Le modèle peut utiliser différentes combinaisons de fonctions d'analyse, et notamment :

- *Analyse en base de données* - Fonctions hautement performantes d'analyse en base de données, utilisant les traitements parallèles pour les opérations de transformation, de calcul mathématique, de géomatique, d'analyse prédictive, de statistique, de séries temporelles et de data mining.
- *Outils d'analyse proposés par des partenaires* - Fonctions d'analyse éprouvées, pour des traitements spécialisés, et exploitables dans le cadre du développement d'applicatifs.

La prédiction de résultats futurs ayant été effectuée au moyen d'un modèle, celui-ci peut être utilisé pour automatiser des décisions optimales en conjuguant localisation et modélisation prédictive. L'intégration d'un processus d'optimisation permet à l'application de résoudre des problèmes métier difficiles, complexes et simultanés, comme par exemple :

- Minimiser simultanément des coûts d'expédition et de stockage,
- Minimiser les dépenses d'investissement tout en limitant les risques liés à une gamme complète,
- Maximiser le chiffre d'affaires et la fidélisation des clients,
- Minimiser les budgets marketing tout en maximisant la génération de leads.

Ces traitements sont effectués tout en prenant en compte différentes contraintes réelles (budgets, embauche de personnel, préférences et comportements des clients, préférences de l'entreprise, entre autres).

## Évaluation par modèle

La construction d'un modèle est suivie de son application aux volumes de données massifs généralement associés à un modèle prédictif. Ce processus constitue une évaluation. Les évaluations consistent, par exemple, à appliquer le modèle à des millions d'enregistrements pour :

- Détecter des activités frauduleuses,
- Calculer des scores de solvabilité (crédit),
- Réévaluer les tarifs concernant des comptes clients,
- Prédire des achats futurs de clients,
- Décider des cibles pour l'envoi de différentes promotions marketing,
- Recommander des offres de crédit à augmenter/diminuer,
- Déterminer si des comptes en défaut de paiement doivent être liquidés,
- Déterminer des stratégies de garantie sur des produits en fonction des appels reçus par le service de maintenance,
- Déterminer les promotions client les plus adaptées pour augmenter les ventes.

## Développement d'algorithmes

Un développeur d'algorithmes dispose d'un choix étendu d'outils de développement, grâce à un environnement intégré de développement utilisant une interface graphique, pour le développement ou l'utilisation de langages de programmation. IBM propose un module complémentaire (plug-in) pour la plateforme Eclipse, environnement de développement intégré « open source ». En outre, les partenaires IBM proposent des environnements de développement intégrés destinés à accélérer le cycle de développement d'algorithmes. IBM propose des ressources pour la plateforme Revolution Analytics R Programming Environment, destinée au développement d'algorithmes, et utilisant le langage R. IBM permet l'accès à différents langages de programmation pour le développement d'algorithmes, notamment les environnements Java, MapReduce, Perl, Python, Lua, R<sup>1</sup>, Fortran, C et C++. En outre, IBM reconnaît les extensions définies par l'utilisateur suivantes :

- *UDF* - Fonctions définies par l'utilisateur, permettant de traiter une rangée.
- *UDA* - Agrégats définis par l'utilisateur, permettant de traiter un groupe de rangées.
- *UDTF* - Fonctions de table définies par l'utilisateur, renvoyant une table, quelle que soit sa taille.
- *UDAP* - Processus d'analyse défini par l'utilisateur, doté d'une flexibilité totale de mise en œuvre d'une ou plusieurs étapes de traitement.
- *Procédures stockées* - Utilisées pour regrouper, sous forme de procédures, des fonctions, agrégats et autres éléments logiques.

Après développement et test, un algorithme est enregistré sur l'appliance. Il bénéficie ainsi de la même accessibilité qu'une autre fonction d'analyse en base de données par le biais des langages de programmation compatibles.

## Récapitulatif

L'appliance PureData System for Analytics, associée à l'environnement IBM Netezza Analytics, constitue une plateforme d'analyse avancée, puissante et extensible qui apporte une simplification du développement, du déploiement et de l'utilisation des outils analytiques, tout en assurant des performances et une évolutivité inégalées. Vous pouvez ainsi conjuguer et exploiter vos propres fonctions d'analyse avec celles proposées par IBM ou par un partenaire tiers, en y accédant par les langages SQL, Java, MapReduce, Perl, Python, Lua, R<sup>1</sup>, Fortran, C ou C++. Votre entreprise bénéficie ainsi de résultats susceptibles de changer la donne grâce à l'architecture parallèle hautes performances de l'appliance d'entrepôt de données IBM Netezza.

## Découvrez par vous-même les capacités de la solution avec notre offre d'essai gratuite

Le programme IBM PureExperience est conçu pour les entreprises qui souhaitent tester la solution PureData System for Analytics. Accessible gratuitement, ce programme permet de tester le système avec vos propres données. Il comprend l'installation sur site, la démonstration de la valeur métier créée, les prestations de formation et de migration des données, l'utilisation du système pour une période déterminée et une ligne d'assistance technique unique. Pour en savoir plus sur le programme et découvrir les possibilités offertes dans votre zone géographique, visitez le site : [ibm.com/PureExperience](http://ibm.com/PureExperience), ou contactez votre interlocuteur IBM.

## La solution IBM PureData System for Analytics en bref

L'offre IBM PureData System for Analytics, associée à la technologie Netezza, intègre une base de données, un serveur et des supports de stockage dans une appliance unique et facile à administrer. La solution ne nécessite qu'une configuration et une administration minimales, et ce, tout en assurant des performances d'analyse accrues et constantes. Elle simplifie considérablement les processus d'analyse métier en consolidant l'ensemble des activités analytiques dans l'appliance, c'est-à-dire là où se trouvent les données, assurant ainsi un niveau de performance parmi les meilleurs de l'industrie. Visitez le site : [ibm.com/PureData](http://ibm.com/PureData) et découvrez comment notre gamme de solutions « Expert Integrated Systems » permet d'éliminer la complexité dans toutes les phases du processus et contribue à créer vraiment de la valeur métier au sein de votre organisation.

## Les solutions IBM Data Warehousing and Analytics en bref

IBM propose la gamme la plus large et complète de logiciels, de matériels et de solutions d'entrepôts de données, de gestion de l'information et d'analyse métier en apportant à ses clients les moyens de maximiser la valeur de leurs actifs en information et de mettre en évidence de nouvelles connaissances pour contribuer à un processus décisionnel plus efficace et rapide et à l'optimisation des résultats de l'entreprise.

### Pour de plus amples informations

Apportez à votre fonction Informatique les moyens de devenir le centre stratégique de votre activité et bénéficiez d'une expertise éprouvée pour assurer votre leadership. Pour en savoir plus sur la gamme PureSystems et plus particulièrement sur l'offre PureData System for Analytics, contactez votre interlocuteur IBM, ou votre partenaire commercial IBM, ou visitez le site web suivant : [ibm.com/PureSystems/PureData](http://ibm.com/PureSystems/PureData).

IBM Global Financing (IGF) est à vos côtés pour l'acquisition des solutions informatiques dont votre entreprise a besoin, et ce, de la manière la plus économique et la plus stratégique possible. IGF vous propose de personnaliser une solution de financement, après examen de votre dossier, de répondre à vos objectifs métier et de développement, de permettre une gestion efficace de votre trésorerie et d'améliorer votre coût total de possession (TCO). Avec IBM Global Financing, vous pouvez financer vos investissements critiques et propulser votre activité vers l'avenir. Pour en savoir plus, visitez le site : [ibm.com/financing/fr](http://ibm.com/financing/fr).



---

**IBM France**

17 Avenue de l'Europe  
92275 Bois Colombes Cedex

IBM, le logo IBM, ibm.com, BigInsights, InfoSphere, PureData, PureExperience et SPSS 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 d'IBM est disponible sur le web sous la rubrique « Copyright and trademark information » à l'adresse [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Netezza est une marque déposée d'IBM International Group B.V., une entreprise IBM.

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

Ces informations concernent les produits, programmes et services commercialisés par IBM France et n'impliquent aucunement l'intention d'IBM de les commercialiser dans d'autres pays. Les références aux produits, programmes et services IBM n'impliquent pas que seuls ces produits, programmes et services peuvent être utilisés. Tout produit, programme ou service équivalent peut être utilisé.

Cette publication a uniquement un rôle informatif.

Ces informations peuvent faire l'objet de modifications sans préavis. Pour bénéficier des informations les plus récentes sur les produits et les services d'IBM, contactez votre interlocuteur commercial ou votre revendeur IBM.

IBM ne donne aucun avis juridique, comptable ou d'audit financier et ne garantit pas que ses produits ou services soient conformes aux lois applicables. Il incombe aux clients de s'assurer que la législation et la réglementation applicables en matière de titres sont respectées, notamment au niveau national.

<sup>1</sup> Revolution R Enterprise for IBM PureData System for Analytics

© Copyright IBM Corporation 2013



Veillez recycler