

IBM SPSS Analytic Server
Version 3.1.1

Présentation

IBM

Important

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations générales figurant à la section «Remarques», à la page 5.

Cette édition s'applique à la version 3.1.1 d'IBM SPSS Analytic Server et à toutes les éditions et modifications ultérieures sauf mention contraire dans les nouvelles éditions.

LE PRESENT DOCUMENT EST LIVRE EN L'ETAT SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFACON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE.

Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. Les informations qui y sont fournies sont susceptibles d'être modifiées avant que les produits décrits ne deviennent eux-mêmes disponibles. En outre, il peut contenir des informations ou des références concernant certains produits, logiciels ou services non annoncés dans ce pays. Cela ne signifie cependant pas qu'ils y seront annoncés.

Pour plus de détails, pour toute demande d'ordre technique, ou pour obtenir des exemplaires de documents IBM, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial.

Vous pouvez également consulter les serveurs Internet suivants :

- <http://www.fr.ibm.com> (serveur IBM en France)
- <http://www.ibm.com/ca/fr> (serveur IBM au Canada)
- <http://www.ibm.com> (serveur IBM aux Etats-Unis)

*Compagnie IBM France
Direction Qualité
17, avenue de l'Europe
92275 Bois-Colombes Cedex*

© Copyright IBM France 2017. Tous droits réservés.

Table des matières

Avis aux lecteurs canadiens	v	Nouveautés de la version 3.1.1	3
Présentation	1	Remarques	5
Architecture	2	Marques	7
Spark et Analytic Server	2		

Avis aux lecteurs canadiens

Le présent document a été traduit en France. Voici les principales différences et particularités dont vous devez tenir compte.

Illustrations

Les illustrations sont fournies à titre d'exemple. Certaines peuvent contenir des données propres à la France.

Terminologie

La terminologie des titres IBM peut différer d'un pays à l'autre. Reportez-vous au tableau ci-dessous, au besoin.

IBM France	IBM Canada
ingénieur commercial	représentant
agence commerciale	succursale
ingénieur technico-commercial	informaticien
inspecteur	technicien du matériel

Claviers

Les lettres sont disposées différemment : le clavier français est de type AZERTY, et le clavier français-canadien de type QWERTY.








OS/2 et Windows - Paramètres canadiens

Au Canada, on utilise :

- les pages de codes 850 (multilingue) et 863 (français-canadien),
- le code pays 002,
- le code clavier CF.

Nomenclature

Les touches présentées dans le tableau d'équivalence suivant sont libellées différemment selon qu'il s'agit du clavier de la France, du clavier du Canada ou du clavier des États-Unis. Reportez-vous à ce tableau pour faire correspondre les touches françaises figurant dans le présent document aux touches de votre clavier.

France	Canada	Etats-Unis
 (Post)		Home
Fin	Fin	End
 (PgAr)		PgUp
 (PgAv)		PgDn
Inser	Inser	Ins
Suppr	Suppr	Del
Echap	Echap	Esc
Attn	Intrp	Break
Impr écran	ImpEc	PrtSc
Verr num	Num	Num Lock
Arrêt défil	Défil	Scroll Lock
 (Verr maj)	FixMaj	Caps Lock
AltGr	AltCar	Alt (à droite)

Brevets

Il est possible qu'IBM détienne des brevets ou qu'elle ait déposé des demandes de brevets portant sur certains sujets abordés dans ce document. Le fait qu'IBM vous fournisse le présent document ne signifie pas qu'elle vous accorde un permis d'utilisation de ces brevets. Vous pouvez envoyer, par écrit, vos demandes de renseignements relatives aux permis d'utilisation au directeur général des relations commerciales d'IBM, 3600 Steeles Avenue East, Markham, Ontario, L3R 9Z7.

Assistance téléphonique

Si vous avez besoin d'assistance ou si vous voulez commander du matériel, des logiciels et des publications IBM, contactez IBM direct au 1 800 465-1234.

Présentation

IBM® SPSS Analytic Server est une solution pour les analyses big data qui combine la technologie IBM SPSS avec les systèmes big data et vous permet de travailler avec les interfaces utilisateur IBM SPSS qui vous sont familières afin de résoudre des problèmes à une échelle jusqu'ici inégalée.

En quoi les analyses big data sont-elles importantes ?

Les volumes de données collectés par les organisations augmentent de façon exponentielle. Par exemple, les sociétés financières et de distribution conservent toutes les transactions de leurs clients pendant un an (ou deux ans, ou dix ans), les opérateurs de télécommunication gardent les enregistrements des détails des appels (CDR) et les lectures des capteurs de terminaux et les sociétés internet stockent tous les résultats des explorations sur le Web.

L'analyse big data est nécessaire lorsqu'il existe :

- Un volume important de données (teraoctets, petaoctets, exaoctets), en particulier lorsqu'il s'agit d'une combinaison de données structurées et non structurées
- Des données s'accumulant/changeant rapidement

Les analyses big data sont également utiles lorsque :

- Un grand nombre (plusieurs milliers) de modèles est généré
- Les modèles sont générés/actualisés de façon fréquente

Les défis

Les organisations qui collectent des volumes massifs de données ont souvent des difficultés à les utiliser, et ce pour plusieurs raisons :

- L'architecture des produits d'analyse traditionnels n'est pas adaptée au calcul distribué, et
- Les algorithmes de statistiques existants ne sont pas conçus pour travailler avec des volumes massifs de données (ces algorithmes attendent que les données viennent vers eux, mais le déplacement de tels volumes de données est trop coûteux), et par conséquent
- Réaliser des analyses d'avant garde sur des volumes massifs de données requiert des compétences nouvelles et une connaissance étroite des systèmes big data. Très peu d'analystes possèdent ces compétences.
- Les solutions en mémoire fonctionnent pour les problèmes de taille moyenne mais ne s'adaptent pas bien aux volumes de données réellement importants.

La solution

Analytic Server offre :

- Une architecture centrée sur les données optimisant les systèmes big data comme Hadoop Map/Reduce avec données dans HDFS.
- Une interface définie pour incorporer de nouveaux algorithmes de statistiques conçus pour aller vers les données.
- Des interfaces utilisateur IBM SPSS familières qui masquent les détails des environnements big data afin que les analystes puissent se concentrer sur l'analyse des données.
- Une solution capable de s'adapter à des problèmes de n'importe quelle taille.

Architecture

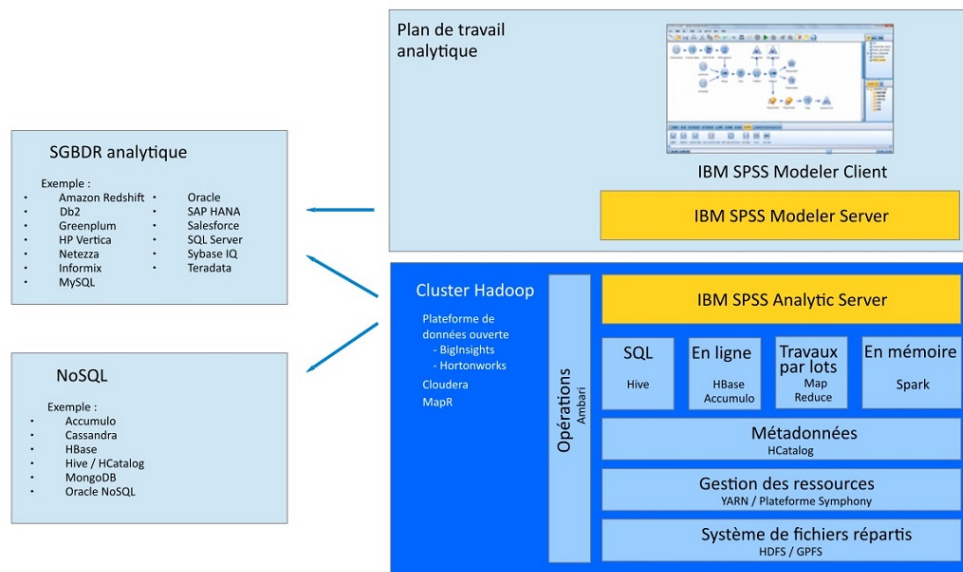


Figure 1. Architecture

Analytic Server se trouve entre une application client et un cloud Hadoop. En supposant que les données résident dans le cloud, la procédure générale employée pour travailler avec Analytic Server est la suivante :

1. Définir des sources de données Analytic Server sur les données du cloud.
2. Définir l'analyse que vous souhaitez réaliser dans l'application client. Pour la version actuelle, l'application client est IBM SPSS Modeler.
3. Lorsque vous lancez l'analyse, l'application client envoie une demande d'exécution Analytic Server.
4. Analytic Server orchestre le travail à exécuter dans le cloud Hadoop et communique les résultats à l'application client.
5. Vous pouvez utiliser les résultats pour définir d'autres analyses et le cycle reprend.

Spark et Analytic Server

Analytic Server s'intègre à Apache Spark pour l'amélioration des performances.

Utilisation ou non de Spark

Si Spark est installé en tant que service Ambari dans le cluster Hadoop, Analytic Server l'utilise pour traiter les travaux relatifs aux big data. Les instructions ci-dessous permettent de déterminer quand Spark n'est pas utilisé.

1. Si la taille de l'ensemble de données est inférieure à 128 Mo, Analytic Server utilise la fonction MapReduce imbriquée dans la machine virtuelle Java d'Analytic Server et n'utilise pas Spark ou le cluster Hadoop.
2. Si Spark n'est pas installé dans le cluster, Analytic Server utilise MapReduce version 2.
3. Analytic Server utilise MapReduce version 2 pour générer des modèles PSM. Lorsqu'un travail se termine par une génération de modèle PSM, Analytic Server utilise Spark pour traiter le travail et effectuer toutes les étapes menant à la génération de modèle, puis pour procéder à l'écriture sur le

disque, et pour utiliser MapReduce afin de générer le modèle PSM. Par exemple, si un travail inclut une jointure suivie d'une génération de modèle PSM, la jointure s'exécute dans Spark et PSM s'exécute sur les données jointes dans MapReduce.

Mode d'utilisation de Spark

Lorsque le service Analytic Server est démarré et détecte que Spark est disponible, il initialise un travail Spark Hadoop qui permet la communication entre les tâches réparties dans le cluster. Ce travail s'exécute tant que le service Analytic Server s'exécute et est utilisé pour toutes les exécutions d'Analytic Server. Cette approche améliore les performances relatives à l'orchestration de plusieurs travaux MapReduce Hadoop car elle élimine la surcharge liée au rechargement de tous les composants Analytic Server pour chaque travail Hadoop.

Spark peut exécuter des travaux MapReduce. Ainsi, Analytic Server peut utiliser des algorithmes Spark "natifs" tels que la jointure, le tri et l'union, s'ils sont disponibles. Simultanément, Analytic Server peut exécuter des algorithmes SPSS Map and Reduce existants dans Spark, sans utiliser directement l'API Hadoop.

Nouveautés de la version 3.1.1

Version 3.1.1

Plateforme

- Prise en charge de Cloudera 5.11 et 5.12
- Prise en charge d'Ubuntu Linux 16.04 (avec Hortonworks Data Platform 2.6 et Cloudera 5.11)
- Cloudera 5.8 et 5.9 ne sont plus pris en charge
- Big Insights 4.1, 4.2 et 4.2.5 ne sont plus pris en charge
- MapR 5.0 n'est plus pris en charge

Sources de données

- Prise en charge d'Apache Hive 2.1.
- MongoDB 2.6 n'est plus pris en charge
- MySQL 5.1 n'est plus pris en charge

Améliorations des performances

- Une procédure d'installation hors ligne plus automatisée est désormais disponible pour HDP. Pour plus d'informations, voir Installation hors ligne
- Prise en charge de clusters multiples. La fonctionnalité de prise en charge de clusters multiples est un approfondissement des capacités de haute disponibilité d'IBM SPSS Analytic Server et fournit un isolement amélioré dans les environnements à plusieurs titulaires. Par défaut, l'installation du service Analytic Server (dans Ambari ou ClouderaManager) entraîne la définition d'un serveur analytique unique.
- Vous pouvez maintenant configurer des files d'attente YARN distinctes pour chaque titulaire Analytic Server. Pour plus d'informations, voir la rubrique portant sur la configuration de files d'attente YARN distinctes pour chaque titulaire Analytic Server pour HDP ou pour Cloudera.
- La conversion SQL des échantillonnages est désormais prise en charge.
- Les améliorations suivantes sont disponibles pour Spark ML :
 - Prise en charge des K-moyennes bissectrices
 - Prise en charge de XGBoost
 - Prise en charge de la régression isotone
- Le croisement des sources de données RDD de Spark peut maintenant être partagé avec différents travaux ASL.

- La section Meilleures pratiques et recommandations pour IBM SPSS Analytic Server est enrichie avec des informations spécifiques à Cloudera.

Pour connaître les informations les plus récentes sur la configuration système requise, reportez-vous aux rapports Detailed System Requirements sur le site du Support technique IBM : <http://publib.boulder.ibm.com/infocenter/prodguid/v1r0/clarity/softwareReqsForProduct.html>. Sur cette page :

1. Entrez SPSS Analytic Server comme nom de produit et cliquez sur **Search**.
2. Sélectionnez la version de votre choix et la portée du rapport, puis cliquez sur **Submit**.

Remarques

Le présent document a été développé pour des produits et des services proposés aux Etats-Unis. Il peut être disponible dans d'autres langues auprès d'IBM. Toutefois, il peut être nécessaire de posséder une copie du produit ou de la version du produit dans cette langue pour pouvoir y accéder.

Le présent document peut contenir des informations ou des références concernant certains produits, logiciels ou services IBM non annoncés dans ce pays. Pour plus de détails, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial IBM. Toute référence à un produit, logiciel ou service IBM n'implique pas que seul ce produit, logiciel ou service puisse être utilisé. Tout autre élément fonctionnellement équivalent peut être utilisé, s'il n'enfreint aucun droit d'IBM. Il est de la responsabilité de l'utilisateur d'évaluer et de vérifier lui-même les installations et applications réalisées avec des produits, logiciels ou services non expressément référencés par IBM.

IBM peut détenir des brevets ou des demandes de brevet couvrant les produits mentionnés dans le présent document. La remise de ce document ne vous donne aucun droit de licence sur ces brevets ou demandes de brevet. Si vous désirez recevoir des informations concernant l'acquisition de licences, veuillez en faire la demande par écrit à l'adresse suivante :

*IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
U.S.A.*

Pour le Canada, veuillez adresser votre courrier à :

*IBM Director of Commercial Relations
IBM Canada Ltd.
3600 Steeles Avenue East
Markham, Ontario
L3R 9Z7 Canada*

Les informations sur les licences concernant les produits utilisant un jeu de caractères double octet peuvent être obtenues par écrit à l'adresse suivante :

*Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan*

LE PRESENT DOCUMENT EST LIVRE EN L'ETAT SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFACON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE. Certaines juridictions n'autorisent pas l'exclusion des garanties tacites, auquel cas l'exclusion ci-dessus ne vous sera pas applicable.

Le présent document peut contenir des inexactitudes ou des coquilles. Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. IBM peut, à tout moment et sans préavis, modifier les produits et logiciels décrits dans ce document.

Les références à des sites Web non IBM sont fournies à titre d'information uniquement et n'impliquent en aucun cas une adhésion aux données qu'ils contiennent. Les documents sur ces sites web ne font pas partie des documents de ce produit IBM et l'utilisation de ces sites web se fait à vos propres risques.

IBM pourra utiliser ou diffuser, de toute manière qu'elle jugera appropriée et sans aucune obligation de sa part, tout ou partie des informations qui lui seront fournies.

Les licenciés souhaitant obtenir des informations permettant : (i) l'échange des données entre des logiciels créés de façon indépendante et d'autres logiciels (dont celui-ci), et (ii) l'utilisation mutuelle des données ainsi échangées, doivent adresser leur demande à :

*IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
U.S.A.*

Ces informations peuvent être soumises à des conditions particulières, prévoyant notamment le paiement d'une redevance.

Le logiciel sous licence décrit dans ce document et tous les éléments sous licence disponibles s'y rapportant sont fournis par IBM conformément aux dispositions de l'ICA, des Conditions internationales d'utilisation des logiciels IBM ou de tout autre accord équivalent.

Les données de performance et les exemples client ne sont présentés qu'à des fins d'illustration. Les performances réelles peuvent varier en fonction des configurations et des conditions d'exploitation.

Les informations concernant des produits non IBM ont été obtenues auprès des fournisseurs de ces produits, par l'intermédiaire d'annonces publiques ou via d'autres sources disponibles. IBM n'a pas testé ces produits et ne peut confirmer l'exactitude de leurs performances ni leur compatibilité. Elle ne peut recevoir aucune réclamation concernant des produits non IBM. Toute question concernant les performances de produits non IBM doit être adressée aux fournisseurs de ces produits.

Les instructions relatives aux intentions d'IBM pour ses opérations à venir sont susceptibles d'être modifiées ou annulées sans préavis, et doivent être considérées uniquement comme un objectif.

Tous les tarifs indiqués sont les prix de vente actuels suggérés par IBM et sont susceptibles d'être modifiés sans préavis. Les tarifs appliqués peuvent varier selon les revendeurs.

Ces informations sont fournies uniquement à titre de planification. Elles sont susceptibles d'être modifiées avant la mise à disposition des produits décrits.

Le présent document peut contenir des exemples de données et de rapports utilisés couramment dans l'environnement professionnel. Ces exemples mentionnent des noms fictifs de personnes, de sociétés, de marques ou de produits à des fins illustratives ou explicatives uniquement. Tous ces noms sont fictifs, et toute ressemblance avec des noms de personnes ou de sociétés réelles serait purement fortuite.

LICENCE DE COPYRIGHT :

Le présent document peut contenir des exemples de données et de rapports utilisés couramment dans l'environnement professionnel. Ces exemples mentionnent des noms fictifs de personnes, de sociétés, de marques ou de produits à des fins illustratives ou explicatives uniquement. Tous ces noms sont fictifs, et toute ressemblance avec des noms de personnes ou de sociétés réelles serait purement fortuite.

Toute copie totale ou partielle de ces programmes exemples et des oeuvres qui en sont dérivées doit comprendre une notice de copyright, libellée comme suit :

© (nom de votre société) (année). Des segments de code sont dérivés des exemples de programmes d'IBM Corp.

Marques

IBM, le logo IBM et ibm.com sont des marques d'International Business Machines Corp. dans de nombreux pays. Les autres noms de produits et de services peuvent être des marques d'IBM ou appartenir à des tiers. La liste actualisée de toutes les marques d'IBM est disponible sur la page Web "Copyright and trademark information" à l'adresse www.ibm.com/legal/copytrade.shtml.

Adobe, le logo Adobe, PostScript et le logo PostScript sont des marques d'Adobe Systems Incorporated aux Etats-Unis et/ou dans certains autres pays.

IT Infrastructure Library est une marque de The Central Computer and Telecommunications Agency qui fait désormais partie de The Office of Government Commerce.

Intel, le logo Intel, Intel Inside, le logo Intel Inside, Intel Centrino, le logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium, et Pentium sont des marques d'Intel Corporation ou de ses filiales aux Etats-Unis et dans certains autres pays.

Linux est une marque de Linus Torvalds aux Etats-Unis et/ou dans certains autres pays.

Microsoft, Windows, Windows NT et le logo Windows sont des marques de Microsoft Corporation aux Etats-Unis et/ou dans certains autres pays.

ITIL est une marque de The Minister for the Cabinet Office et est enregistrée au bureau américain Patent and Trademark Office.

UNIX est une marque enregistrée de The Open Group aux Etats-Unis et/ou dans certains autres pays.

Cell Broadband Engine est une marque de Sony Computer Entertainment, Inc., aux Etats-Unis et/ou dans certains autres pays, et est utilisée sous license.

Linear Tape-Open, LTO, le logo LTO, Ultrium et le logo Ultrium sont des marques de HP, IBM Corp. et Quantum aux Etats-Unis et/ou dans certains autres pays.

