

IBM SPSS Analytic Server
Version 2

Guide d'installation et de configuration

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 31.

Cette édition s'applique à la version 2.0.0 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 2015. Tous droits réservés.

Table des matières

Avis aux lecteurs canadiens	v	Chapitre 4. Migration.	19
Chapitre 1. Présentation	1	Utilisation de l'outil de migration	20
Architecture	2	Chapitre 5. Apport de modifications à	
Nouveautés pour l'installation dans la version 2	2	votre système.	23
Chapitre 2. Installation	5	Utilisation de l'outil d'administration	23
Prérequis	5	Modification manuelle des propriétés de	
Prérequis pour cluster à haute disponibilité	5	configuration	24
Installation	6	Déploiement sur de nouveaux membres du cluster à	
Chapitre 3. Configuration	9	haute disponibilité	26
Utilisation de l'outil de configuration	9	Chapitre 6. Désinstallation	27
Création d'un registre d'utilisateurs	12	Chapitre 7. Traitement des incidents	29
Activation des sources de données HCatalog	13	Remarques	31
Activation de la prise en charge d'Essentials for R	15	Marques	33
Configuration d'IBM SPSS Modeler pour son			
utilisation avec IBM SPSS Analytic Server	16		

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
 (Pos1)		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.

Chapitre 1. Présentation

IBM® SPSS Analytic Server est une solution pour les analyses de volumes massifs de données (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

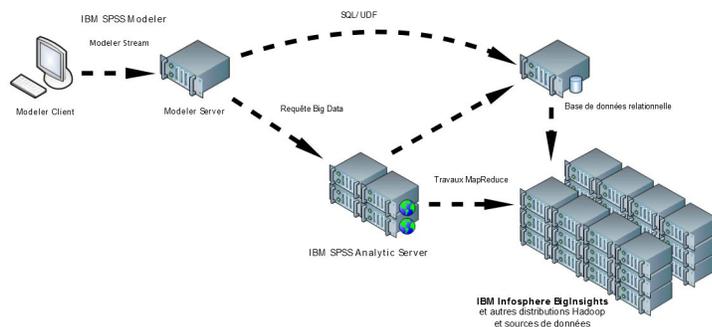


Figure 1. Architecture

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

1. Définir des sources de données Analytic Server sur les données du nuage.
2. Définir l'analyse que vous souhaitez réaliser dans l'application client. Pour l'édition 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. L'Analytic Server orchestre le travail à exécuter dans le nuage Hadoop et communique les résultats à l'application client.
5. Vous pouvez utiliser les résultats pour définir d'autres analyses et le cycle reprend.

Nouveautés pour l'installation dans la version 2

Installation et configuration

L'installation du produit utilise dorénavant IBM Installation Manager pour permettre une installation transparente lorsque vous installez plusieurs produits IBM.

Les étapes de configuration d'Analytic Server sont traitées dans un outil distinct qui est exécuté une fois que toutes les extensions du produit ont été installées.

Plateforme

Prise en charge de nouveaux systèmes d'exploitation et de distributions Hadoop supplémentaires.

Systèmes d'exploitation

Analytic Server fonctionne maintenant Microsoft Windows.

Distributions Hadoop

Analytic Server peut être configuré pour fonctionner avec Hortonworks Data Platform sous Microsoft Windows.

Adaptateurs HCatalog

Des sources de données peuvent être définies pour Apache Cassandra, MongoDB et Oracle NoSQL. Des étapes de configuration supplémentaires sont requises pour activer ces sources de données.

Clusters à haute disponibilité

Analytic Server peut être installé et configuré sur un cluster à haute disponibilité via IBM WebSphere Liberty Network Deployment.

Chapitre 2. Installation

Prérequis

Avant d'installer Analytic Server, consultez les informations suivantes.

Configuration système requise

Pour les informations les plus récentes sur la configuration système requise, reportez-vous au document Detailed system 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 voulue et la portée du rapport, puis cliquez sur **Submit**.

Privilèges d'administration

L'utilisateur qui procède à l'installation d'Analytic Server doit disposer d'un accès en écriture au système de fichiers distribué Hadoop (HDFS).

IBM Installation Manager

Une version à jour d'IBM Installation Manager doit être installée et configurée afin d'utiliser un référentiel contenant les fichiers d'installation Analytic Server. Si vous disposez d'une installation plus ancienne d'IBM Installation Manager, vous devez la mettre à niveau vers une version prise en charge, dans le cadre de l'installation.

Si ce programme n'est pas présent sur le système, téléchargez et installez IBM Installation Manager depuis le site du support d'IBM IBM Corp. : <http://www.ibm.com/support>.

Important : Sous Microsoft Windows, installez IBM Installation Manager depuis un compte administrateur. Vous pouvez effectuer une installation de groupe depuis un autre compte utilisateur en spécifiant : `./groupinstc -acceptLicense`. Reportez-vous à la documentation IBM Installation Manager pour plus d'informations : http://www-01.ibm.com/support/knowledgecenter/SSDV2W/im_family_welcome.html

Hive/HCatalog

Si vous comptez utiliser des sources de données NoSQL, configurez Hive et HCatalog pour accès distant. Vérifiez également que le fichier `hive-site.xml` contient une propriété `hive.metastore.uris` sous la forme `thrift://<nom_hôte>:<port>` qui pointe vers le serveur Thrift Hive Metastore actif. Pour plus d'informations, reportez-vous à la documentation de votre distribution Hadoop.

Référentiel de métadonnées

Par défaut, Analytic Server installe et utilise une base de données Apache Derby. Vous avez toutefois la possibilité de configurer Analytic Server afin d'utiliser une installation DB2 existante. La base de données doit être codée en UTF-8. IBM DB2 10.5 Workgroup Server Edition est disponible sur le média Analytic Server. Pour plus d'informations, voir : http://www-01.ibm.com/support/knowledgecenter/SSEPGG_10.5.0/com.ibm.db2.luw.kc.doc/welcome.html.

Cluster à haute disponibilité

Voir «Prérequis pour cluster à haute disponibilité» pour examiner les prérequis pour l'installation de Analytic Server sur un cluster à haute disponibilité.

Prérequis pour cluster à haute disponibilité

Vous devez respecter certains prérequis supplémentaires pour installer Analytic Server sur un cluster à haute disponibilité.

Systèmes et configuration DNS

Les systèmes utilisés dans l'environnement doivent pouvoir communiquer entre eux. Seuls les systèmes appartenant à un domaine doivent avoir le suffixe du domaine. Par exemple, seuls les systèmes enregistrés avec un domaine "IBM.COM" doivent contenir un suffixe "ibm.com", dans le nom de domaine complet (FQDN, Fully Qualified Domain Name).

Si un système n'est pas enregistré dans le domaine IBM.COM mais possède un suffixe FQDN avec "ibm.com", alors vous pouvez saisir ce FQDN dans le serveur DNS, avec les systèmes dotés de ce type de serveurs dans la configuration TCP. Autrement, les FQDN de tous les systèmes concernés doivent être indiqués dans les fichiers d'hôtes de tous les systèmes impliqués.

Référentiel de métadonnées

Analytic Server doit utiliser DB2 s'il est installé sur un cluster à haute disponibilité.

Zookeeper 3.4.5

Trois instances Zookeeper au minimum sont requises. Reportez-vous à la documentation d'installation et de configuration de Zookeeper : http://zookeeper.apache.org/doc/r3.4.5/zookeeperStarted.html#sc_RunningReplicatedZooKeeper.

Remarque : Ces étapes doivent être réalisées sur chaque système sur lequel Zookeeper est installé.

Liberty Network Deployment 8.5.5.4

L'environnement à haute disponibilité se compose d'une collectivité Liberty avec un contrôleur, deux serveurs secondaires et au moins deux membres de cluster. Reportez-vous à la documentation de WebSphere Application Server Network Deployment pour plus d'informations sur la configuration d'une collectivité Liberty : http://www-01.ibm.com/support/knowledgecenter/SSAW57/mapfiles/product_welcome_wasnd.html.

Équilibreur de charge

Votre cluster à haute disponibilité doit disposer d'un équilibreur de charge prenant en charge l'affinité de session. Analytic Server identifie les sessions avec le cookie "request-token". Celui-ci identifie une session pour la durée de connexion d'un utilisateur et son utilisation dans une affinité de session contrôlée par l'application. Consultez la documentation de votre équilibreur de charge spécifique pour plus d'informations sur sa prise en charge de l'affinité de session.

Remarque : Le noeud leader ne doit pas être inclus dans l'équilibreur de charge du cluster. Le leader est le tout premier noeud que vous démarrez.

Visual C++ 2008 (Windows uniquement)

Si vous disposez d'un membre de cluster exécutant Analytic Server sur Windows, vous devrez installer le contexte d'exécution Visual C 2008 avant d'installer Analytic Server.

Installation

Les fichiers d'application Analytic Server sont installés sur le système hôte à l'aide d'IBM Installation Manager. Ces fichiers d'installation sont disponibles sur le support physique ou peuvent être téléchargés depuis IBM Passport Advantage.

1. Lancez IBM Installation Manager :

Mode interface graphique :

```
{répertoire_installation_IBM Installation Manager}/eclipse/IBMIM
```

Mode de ligne de commande :

```
{répertoire_installation_IBM Installation Manager}/eclipse/tools/imcl -c
```

Remarque : L'utilisateur exécutant IBM Installation Manager est celui qui exécute le processus Analytic Server. Cet utilisateur doit pouvoir accéder à l'environnement Hadoop.

2. Si le répertoire d'installation n'est pas configuré, spécifiez le chemin du référentiel. Par exemple, spécifiez le disque d'installation monté ou un emplacement sur le système de fichiers hôte, le réseau ou une adresse HTTP.
3. Dans le menu principal, sélectionnez **Installer**.
4. Sélectionnez le package à installer. L'option disponible dépend de la version Analytic Server que vous avez acquise.

IBM SPSS Analytic Server

Ce package est compatible avec la plupart des systèmes.

IBM SPSS Analytic Server High Availability

Ce package est requis si vous comptez exécuter Analytic Server dans un cluster à haute disponibilité. Les différences de configuration of IBM SPSS Analytic Server High Availability sont documentées à la place.

5. Lisez le contrat de licence et acceptez ses dispositions.
6. Spécifiez le groupe de package et le répertoire d'installation.
 - Un nouveau groupe de package est requis pour l'installation de IBM SPSS Analytic Server.
 - Spécifiez le répertoire d'installation des ressources partagées. Vous ne pouvez spécifier le répertoire de ressources partagées que la première fois que vous installez un package.
7. Sélectionnez les fonctionnalités à installer.
8. Consultez le récapitulatif des informations et procédez à l'installation. Les fichiers d'application sont installés dans le répertoire indiqué.
9. Une configuration supplémentaire est nécessaire. Pour plus d'informations quant à une nouvelle configuration, reportez-vous à la rubrique : «Utilisation de l'outil de configuration», à la page 9. Pour plus d'informations sur la migration d'une configuration existante vers une nouvelle installation, reportez-vous à la rubrique : Chapitre 4, «Migration», à la page 19. Si vous exécutez IBM Installation Manager en mode d'interface graphique, l'option de lancement de l'outil de configuration ou de migration vous est présentée lors de la dernière étape.

Si des problèmes surviennent lors de l'installation, vous pourrez les corriger à l'aide des journaux d'IBM Installation Manager. Vous pouvez accéder à ces journaux depuis le menu principal d'IBM Installation Manager.

Vous pouvez exécuter IBM Installation Manager en mode silencieux afin d'automatiser l'installation. Pour plus d'informations, reportez-vous à la documentation IBM Installation Manager : http://www-01.ibm.com/support/knowledgecenter/SSDV2W/im_family_welcome.html.

Chapitre 3. Configuration

Après l'installation, vous devez configurer Analytic Server avant d'utiliser le produit.

Remarque : Les conventions suivantes sont utilisées pour les chemins de fichier Analytic Server.

- {RACINE_AS} se réfère à l'emplacement dans lequel Analytic Server est déployée ; par exemple, /opt/IBM/SPSS/AnalyticServer/{version}.
- {RACINE_SERVEUR_AS} se réfère à l'emplacement des fichiers de configuration, des journaux et du serveur. Par exemple, /opt/IBM/SPSS/AnalyticServer/{version}/ae_wlpserver/usr/servers/aeserver.
- {DOSSIER_PRINCIPAL_AS} désigne l'emplacement sur HDFS utilisé par Analytic Server en tant que dossier principal.

Etapes de configuration

1. [Facultatif] Suivez les étapes de la rubrique Activation des sources de données HCatalog si vous désirez exécuter Analytic Server sur des sources de données NoSQL ou XML.

Remarque : Vous pouvez effectuer une configuration simple d'Analytic Server dès maintenant et activer plus tard des sources de données HCatalog à l'aide de l'outil d'administration.

2. Exécutez l'outil de configuration. L'outil automatise un certain nombre d'étapes qui devraient autrement être réalisées manuellement. Si vous avez utilisé la version de l'interface graphique utilisateur d'IBM Installation Manager, l'outil de configuration se lance automatiquement. Sinon, vous devez exécuter le script **configtool** dans {RACINE_AS}/bin.
3. Créez un registre d'utilisateurs pouvant être ajoutés au système Analytic Server.
4. Configurez IBM SPSS Modeler pour son utilisation avec Analytic Server.
5. [Facultatif] Actiez la prise en charge de Essentials for R si vous désirez que Analytic Server exécute du code R.
6. Vérifiez que Analytic Server est en opération. Vous pouvez lancer le processus en exécutant le script **start** situé sous {RACINE_AS}/bin.
7. Ouvrez un navigateur Web et entrez l'adresse <http://localhost:8080/analyticserver/admin/ibm>. Cette adresse URL affiche la boîte de dialogue de connexion de la console Analytic Server. Connectez-vous avec les données d'identification de l'administrateur Analytic Server défini dans l'outil de configuration.

Utilisation de l'outil de configuration

L'outil de configuration automatise un certain nombre d'étapes qui devraient autrement être réalisées manuellement. Cet outil doit pouvoir accéder à l'environnement HDFS.

Si vous avez utilisé la version de l'interface graphique utilisateur d'IBM Installation Manager, l'outil de configuration se lance automatiquement. Suivez les invites pour configurer votre système.

Vous pouvez également lancer l'outil de configuration en exécutant le script **configtool** dans {RACINE_AS}\bin. Par défaut, l'outil détecte s'il doit être lancé dans l'interface graphique ou en mode console. Vous pouvez imposer le mode console en lui transmettant l'argument **-c** ou le mode silencieux en lui indiquant le chemin d'un fichier spécifiant la configuration voulue. Voir le fichier **configtool.response** dans {RACINE_AS}\bin fourni comme exemple.

Bienvenue

L'étape Bienvenue détermine le chemin à suivre par l'outil de configuration.

Standard

Le chemin Standard automatise une grande partie des décisions et pose des hypothèses simplificatrices sur votre système. Sélectionnez cette option si vous voulez configurer rapidement un environnement autre que de production. Vous pourrez toujours modifier votre configuration dans l'outil Administration ou migrer votre configuration vers un autre environnement par la suite.

Dans le cas d'une installation standard, l'outil vous invitera à indiquer l'emplacement du serveur namenode de Hadoop, et à réclamer au fur et à mesure et en cas de besoin des informations supplémentaires.

Dans le cas d'une installation standard de haute disponibilité, l'outil de configuration vous réclamera des informations sur votre environnement WebSphere® Application Server Liberty Network Deployment et votre ensemble Zookeeper.

Personnalisée

Permet de réaliser des tâches de configuration plus avancées. Cette option est requise si vous désirez configurer à ce state NoSQL ou des bases de données relationnelles pour leur utilisation avec Analytic Server.

Les sections ci-dessous vous renseignent plus spécifiquement sur le chemin personnalisé.

Cluster Hadoop

Les étapes de cluster Hadoop configurent Analytic Server pour son utilisation avec votre distribution Hadoop. L'outil de configuration peut reconnaître automatiquement votre distribution, ou vous pouvez sélectionner manuellement les fichiers nécessaires.

Collecter automatiquement les fichiers de votre distribution Hadoop

L'outil tentera de détecter la distribution Hadoop et réclamera au fur et à mesure et en cas de besoin des informations supplémentaires.

1. Vous devez toujours indiquer le nom d'hôte du cluster Hadoop.
2. Si votre cluster n'est pas configuré avec SSH et avec l'authentification de clé publique de l'utilisateur exécutant l'outil de configuration, vous devrez soumettre des données d'identification valides pour que l'outil de configuration puisse collecter des fichiers depuis le cluster Hadoop.

Restriction : Seul le chiffrement SSH RSA est pris en charge.

3. Si votre cluster Hadoop a été installé dans un emplacement non standard, vous pouvez être invité à indiquer l'emplacement d'installation.

Collecter manuellement les fichiers de votre distribution Hadoop

Identifiez la distribution Hadoop que vous utilisez. L'outil de configuration identifiera les fichiers que vous avez besoin de collecter depuis le cluster Hadoop. Copiez ces fichiers vers l'emplacement spécifié et poursuivez les étapes de configuration.

Remarque : Le fichier JAR `hadoop-lzo-*.jar` n'est requis que si la valeur `"com.hadoop.compression.lzo.LzoCodec"` est spécifiée pour le paramètre `io.compression.codecs` dans le fichier `core-site.xml`.

Hcatalog/Hive

Les étapes Hcatalog/Hive ne sont affichées que dans le cadre d'une configuration personnalisée.

Se connecter automatiquement et détecter votre configuration HCatalog / Hive

L'outil détectera automatiquement votre version de HCatalog / Hive et collectera les fichiers requis pour configurer Analytic Server.

Sélectionner manuellement votre configuration HCatalog / Hive

Identifiez les bases de données NoSQL que vous utilisez. L'outil de configuration identifiera les fichiers que vous devez collecter. Copiez ces fichiers vers l'emplacement spécifié et poursuivez les étapes de configuration.

Ignorer la configuration HCatalog / Hive

Vous pouvez configurer HCatalog / Hive plus tard.

JDBC

Les étapes JDBC ne sont présentées que dans le cadre d'une configuration personnalisée. Copiez les pilotes de la base de données que vous utilisez vers l'emplacement spécifié et poursuivez les étapes de configuration.

Tableau 1. Bases de données prises en charge

Base de données	Versions prises en charge	Fichiers jar du pilote JDBC	Fournisseur
DB2 for Linux, UNIX, and Windows	11, 10.5, 10.1, 9.7	db2jcc.jar	IBM
DB2 z/OS	11, 10	db2jcc.jar, db2_license_cisuz.jar	IBM
Greenplum	5, 4.2.x	postgresql.jar	Greenplum
Netezza	7, 6.x	nzjdbc.jar	IBM
Oracle	12c, 11g R2 (11.2)	ojdbc6.jar, ora118n.jar	Oracle
SQL Server	2014, 2012, 2008 R2	sqljdbc4.jar	Microsoft
Sybase IQ	16.x, 15.4, 15.2	jconnect70.jar	Sybase
Teradata	14, 14.1, 15	tdgssconfig.jar, terajdbc4.jar	Teradata

Serveur d'application

Les étapes Serveur d'application dépendent du type d'installation.

- Dans une installation sans haute disponibilité, vous pouvez configurer une communication sécurisée (SSL/TLS) et définir les ports sur lesquels Analytic Server est à l'écoute.
- Dans une installation de haute disponibilité, l'outil de configuration vous réclamera des informations sur votre environnement WebSphere® Application Server Liberty Network Deployment.

Zookeeper

L'étape de configuration Zookeeper n'est requise que dans le cadre d'une installation de haute disponibilité. Indiquez la liste des serveurs dans votre ensemble Zookeeper.

Référentiel de métadonnées

Ces étapes dépendent de votre type d'installation.

- Dans une installation sans haute disponibilité, vous avez le choix entre Derby et DB2. Si vous choisissez DB2, ce référentiel doit être installé et disponible avant d'exécuter cet outil de configuration. Pour plus d'informations, voir «Prérequis», à la page 5.
- Une configuration d'une installation de haute disponibilité doit toujours utiliser DB2 comme référentiel de métadonnées et l'instance DB2 doit être installée et disponible avant d'exécuter l'outil de configuration.

Analytic Server

Identifiez le principal faisant office d'administrateur Analytic Server. Ce principal pourra créer, supprimer et gérer les titulaires.

Récapitulatif

Cette étape vous permet d'examiner les étapes de configuration avant de la finaliser.

Post-configuration

L'étape suivant la configuration indiquera que la configuration a abouti. En cas d'échec, consultez le fichier journal dans le répertoire logs.

Création d'un registre d'utilisateurs

Le titulaire par défaut créé lors de l'installation et de la configuration utilise le profil WebSphere Liberty pour authentifier les utilisateurs et ce dernier peut à son tour référencer un serveur LDAP externe ou définir un registre de base d'utilisateurs et de groupes.

Remarque : Après avoir modifié le fichier {RACINE_SERVEUR_AS}/server.xml en configurant WebSphere Liberty Profile pour authentifier les utilisateurs, ajoutez ces utilisateurs au système Analytic Server. Reportez-vous au manuel *IBM SPSS Analytic Server - Guide d'administration* pour plus d'informations sur la gestion des titulaires.

LDAP

Le registre LDAP vous permet d'authentifier les utilisateurs via un serveur LDAP externe tel que Active Directory ou OpenLDAP. Voici un exemple de registre LDAP pour OpenLDAP.

```
<ldapRegistry
  baseDN="ou=people,dc=aeldap,dc=org"
  ldapType="Custom"
  port="389"
  host="server"
  id="OpenLDAP"
  bindDN="cn=admin,dc=aeldap,dc=org"
  bindPassword="{xor}Dz4sLG5tbGs="
  searchTimeout="300000m"
  recursiveSearch="true">
  <customFilters
    id="customFilters"
    userFilter="(&uid=%v)(objectClass=inetOrgPerson)"
    groupFilter="(&cn=%v)(|(objectclass=organizationalUnit))"
    groupMemberIdMap="posixGroup:memberUid"/>
</ldapRegistry>
```

Pour obtenir d'autres exemples de configurations, consultez le dossier des modèles {RACINE_AS}/ae_wlpserver/templates/config.

De base

Le registre de base vous permet de définir une base de données d'utilisateurs et de groupes dans le fichier {RACINE_SERVEUR_AS}/server.xml. Les mots de passe peuvent être codés afin de brouiller leur valeur via l'outil securityUtility, situé sous le répertoire {RACINE_AS}/ae_wlpserver/bin.

Le registre de base est utile dans un environnement de bac à sable, mais n'est pas recommandé pour un environnement de production.

```

<basicRegistry id="basic" realm="ibm">
  <user name="user1" password="{xor}Dz4sLG5tbGs="/>
  <user name="user2" password="Pass"/>
  <user name="user3" password="Pass"/>
  <user name="user4" password="Pass"/>
  <user name="admin" password="{xor}KzosKw="/>
  <group name="Development">
    <member name="user1"/>
    <member name="user2"/>
  </group>
  <group name="QA">
    <member name="user3"/>
    <member name="user4"/>
  </group>
  <group name="ADMIN">
    <member name="user1"/>
    <member name="admin"/>
  </group>
</basicRegistry>

```

Activation des sources de données HCatalog

Analytic Server prend en charge les données XML et les bases de données NoSQL via Hive/HCatalog.

L'outil de configuration prépare votre installation Analytic Server afin d'utiliser Hive/HCatalog, mais certaines étapes manuelles doivent être réalisées. Vous devez copier plusieurs fichiers JAR dans le répertoire {DOSSIER_PRINCIPAL_HIVE}/auxlib/ étant donné que Analytic Server implémente certaines interfaces requises par HCatalog. Analytic Server ne peut pas utiliser directement de gestionnaires de stockage pour Hive vu que HCatalog utilise une API légèrement différente.

Les sections suivantes recensent les fichiers nécessaires en fonction des sources de données que vous désirez.

Important : Ces fichiers doivent être copiés dans le répertoire {REPERTOIRE_PRINCIPAL_HIVE}/auxlib/ avant de lancer l'outil de configuration ou l'outil d'administration pour que Analytic Server puisse les utiliser.

Important : Hortonworks Data Platform sur Windows ne reconnaît pas le répertoire {REPERTOIRE_PRINCIPAL_HIVE}/auxlib/. Pour activer HCatalog, vous pouvez:

1. Copier les fichiers requis, tels que décrits ci-dessous, vers le répertoire {REPERTOIRE_PRINCIPAL_HIVE}/lib/.
2. Supprimer le répertoire {REPERTOIRE_PRINCIPAL_HIVE}/auxlib/.
3. Mettre à jour les fichiers hive-site.xml et hive-config.cmd en supprimant les références au répertoire {REPERTOIRE_PRINCIPAL_HIVE}/auxlib/.
4. Redémarrer le cluster Hortonworks Data Platform afin que les modifications prennent effet.

Pour plus d'informations, voir <http://hortonworks.com/community/forums/topic/hive-auxlib-issues/>.

Toutes les installations Hive/HCatalog

1. Les fichiers suivants résident dans l'archive hive_auxlib située sous le répertoire {RACINE_AS}/add-ons.

```
com.ibm.spss.com.hcatalogstoragehandler-{version}.jar
```

2. Copiez la version la plus récente de hivexmlserde-{version}.jar depuis le référentiel Maven Central sur le site <http://search.maven.org/#search%7Cga%7C1%7Chivexmlserde>.
3. [Facultatif] Si vous comptez utiliser des types Hive complexes (mappes, tableaux, structures), copiez les fichiers JAR suivants. Ces fichiers sont disponibles sous l'installation HCatalog. Avant Hive 0.11.0,

il s'agit de {DOSSIER_PRINCIPAL_HCATALOG}/share/hcatalog ; à compter de Hive 0.11.0, il s'agit de {DOSSIER_PRINCIPAL_HIVE}/hcatalog/share/hcatalog.

hcatalog-core-{version}.jar (pour hadoop 1.2)
hive-hcatalog-{version}.jar (pour hadoop 2.x)
jackson-core-asl-1.8.8.jar
jackson-mapper-asl-1.8.8.jar

4. Si vous utilisez Hive version 0.14.0, téléchargez le fichier JAR suivant. <https://github.com/dvasilen/Hive-HCatalog-Compatibility/raw/master/release/hive-hcatalog-compatibility-0.14.0.jar>

Données XML

1. [Facultatif] Si vous comptez utiliser des fichiers XML compressés, copiez les fichiers JAR suivants depuis l'archive hive_auxlib située sous le répertoire {RACINE_AS}/add-ons.
com.ibm.spss.com.ibm.biginsights.compression_2.1.jar (distribution BigInsights Hadoop uniquement)
com.ibm.spss.hivexmlserdecomp-2.0.0.0.jar

Apache Accumulo

1. Les fichiers suivants sont disponibles sous l'installation HCatalog. Avant Hive 0.11.0, il s'agit de {DOSSIER_PRINCIPAL_HCATALOG}/share/hcatalog ; à compter de Hive 0.11.0, il s'agit de {DOSSIER_PRINCIPAL_HIVE}/hcatalog/share/hcatalog.
commons-io-{version}.jar
2. Le gestionnaire de stockage Apache Accumulo est disponible sur le site <https://github.com/dvasilen/accumulo-hive-storage-manager/tree/master>. Il s'agit ici de la branche principale ; vous devez ensuite basculer sur celle correspondant à votre version de Accumulo et Hadoop. Le fichier JAR est situé sous le répertoire /release et son nom est similaire à :
accumulo-hive-storage-handler_<version_hadoop>-<version_accumulo>.jar
3. Les fichiers suivants sont disponibles sous l'installation Apache Accumulo.
accumulo-trace-{version}.jar
accumulo-fate-{version}.jar
accumulo-core-{version}.jar
accumulo-server-{version}.jar
accumulo-start-{version}.jar

Apache Cassandra

1. Les fichiers suivants sont disponibles dans l'édition Apache Cassandra 1.2.9 sur le site <http://www.apache.org/dyn/closer.cgi?path=/cassandra/1.2.9/apache-cassandra-1.2.9-bin.tar.gz>
apache-cassandra-1.2.9.jar
apache-cassandra-clientutil-1.2.9.jar
apache-cassandra-thrift-1.2.9.jar
2. Le gestionnaire de stockage Apache Cassandra est disponible sur le site <https://github.com/dvasilen/Hive-Cassandra>. Il s'agit ici de la branche principale, depuis laquelle vous devez bifurquer vers celle correspondant à votre version Cassandra et Hadoop. Le fichier JAR est situé sous le répertoire /release et son nom est similaire à :
hive-<version_hive>-hadoop-<version_hadoop>-cassandra-<version_cassandra>.jar

Remarque : Pour les versions Hadoop antérieures à la version, 2.0, il n'existe pas de ramification hadoop du nom de fichier ou de branche.

Remarque : Analytic Server peut communiquer avec Cassandra 2.0, mais ses fichiers JAR sont compilés pour Java 7. Par conséquent, vous devrez copier les fichiers JAR de Cassandra 1.2.9 JAR vu qu'ils sont compilés pour Java 6.

HBase

Vous n'avez pas besoin de copier de fichiers supplémentaires pour HBase.

MongoDB

1. Les fichiers suivants doivent être générés localement pour votre distribution Hadoop depuis la branche "principale" du projet "mongodb/mongo-hadoop". Voir <https://github.com/mongodb/mongo-hadoop> pour plus de détails.

```
mongo-hadoop-core-{version}.jar  
mongo-hadoop-hive-{version}.jar
```

2. Le fichier suivant est disponible sur le site <https://github.com/downloads/mongodb/mongo-java-driver/mongo-2.10.1.jar>
mongo-2.10.1.jar

Oracle NoSQL

1. Le fichier suivant est disponible sur le site <https://github.com/vilcek/HiveKVStorageHandler2/blob/master/dist/HiveKVStorageHandler.jar>

```
HiveKVStorageHandler.jar
```

2. Le fichier suivant est disponible sous l'installation Apache Hive
avro-*<version>.jar

Activation de la prise en charge d'Essentials for R

Analytic Server prend en charge l'évaluation de modèles R et l'exécution de scripts R.

Restriction : La prise en charge de R est actuellement limitée aux installations d'Analytic Server sous Linux.

Pour configurer la prise en charge de R suite à une installation réussie d'Analytic Server :

1. Installez R Engine sur le serveur qui héberge Analytic Server.

```
wget http://cran.r-project.org/src/base/R-3/R-3.1.0.tar.gz  
tar -xzf R-3.1.0.tar.gz  
cd R-3.1.0  
./configure --enable-R-shlib  
make
```

Important : Installez R Engine dans un emplacement accessible à tous les utilisateurs vu que R s'exécute sur le cluster Hadoop en utilisant un utilisateur différent de l'utilisateur Analytic Server.

Important : Dans une installation à haute disponibilité, R Engine doit être installé sous le même emplacement sur chaque membre du cluster.

2. Installez Essentials for R sur le serveur qui héberge Analytic Server en exécutant le fichier du programme d'installation et en suivant les instructions affichées à l'écran. Vous pouvez télécharger Essentials for R depuis le site <https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=swg-tspssp>.
3. Déployez R Engine et R Component sur Hadoop.
 - a. Si Analytic Server et tous les noeuds Hadoop possèdent la même version de système d'exploitation et la même architecture de processeur :

- 1) Créez une archive pour R Engine à l'aide du script suivant

```
#!/usr/bin/env bash  
echo Creating R.zip...  
cd /tmp  
rm -r -f R  
rm -f R.zip
```

```

mkdir R
export R_HOME=/home/hdpadmin/APPS/R/R-3.1.0
cp -r $R_HOME/* ./R/
cp $(ls /usr/lib64/libgfortran.so.3 | sort -u) ./R/lib/
cp -P /usr/lib64/libgfortran.so.3.0.0 ./R/lib
rm -r ./R/doc ./R/src ./R/include ./R/tests
zip -r R.zip R

```

Remarque : Deux espaces figurent entre `-d\` et `-f3` dans le script. Le document PDF semble contenir deux espaces, mais en fait il s'agit d'un seul espace élargi. Lorsque vous copiez et collez ce script depuis le fichier PDF, insérez le second espace avant d'exécuter le script.

Remarque : Les lignes qui copient les bibliothèques Fortran dépendent de la version Fortran installée. Par exemple, si la version 1.0 est installée, ces lignes indiquent comme suit. Modifiez-les comme requis pour votre système.

```

cp -P /usr/lib64/libgfortran.so.1 ./R/lib
cp /usr/lib64/libgfortran.so.1.0.0 ./R/lib

```

- 2) Copiez l'archive R.zip dans le répertoire `{RACINE_AS}/ae_wlpserver/usr/servers/aeserver/configuration/linux/app_64`.
- b. Si les versions des systèmes d'exploitation pour les noeuds Analytic Server et les noeuds Hadoop sont différentes, installez le moteur R Engine et Essentials for R sur chaque noeud Hadoop. Installez-les dans un répertoire avec le même chemin et le même nom que ceux où le moteur R est installé sur Analytic Server. Si vous installez Essentials for R sur le noeud Hadoop, spécifiez le répertoire où R est installé et sautez l'étape où l'emplacement `../ext_64/bin` est spécifié.
- c. Lancez le script **hdfsUpdate** situé sous le répertoire `{RACINE_AS}/bin` afin de propager les modifications dans HDFS.
4. Sur une installation à haute disponibilité, vous devez également lancer le script **deployRemote** situé sous le répertoire `{RACINE_AS}/bin` afin de propager les modifications aux instances Analytic Server sur chaque membre du cluster.
5. Vous devez également installer Essentials for R sur la machine qui héberge SPSS Modeler Server.

Configuration d'IBM SPSS Modeler pour son utilisation avec IBM SPSS Analytic Server

Pour activer SPSS Modeler pour son utilisation avec Analytic Server, vous devez effectuer certaines mises à jour dans l'installation de SPSS Modeler Server.

1. Configurez SPSS Modeler Server en l'associant à une installation Analytic Server.
 - a. Ouvrez le fichier `options.cfg` situé sous le sous-répertoire `config` du répertoire d'installation racine du serveur et ajoutez ou éditez les lignes suivantes :

```

as_ssl_enabled, {Y|N}
as_host, "{SERVEUR_AS}"
as_port, PORT
as_context_root, "{RACINE_CONTEXTE}"
as_tenant, "{TITULAIRE}"
as_prompt_for_password, {Y|N}

```

as_ssl_enabled

Spécifiez Y si la communication sécurisée est configurée sur Analytic Server ; sinon, spécifiez N.

as_host

Adresse IP du serveur sur lequel réside Analytic Server.

as_port

Port sur lequel Analytic Server est à l'écoute (par défaut, il s'agit du port 8080).

as_context_root

Racine de contexte Analytic Server (par défaut, il s'agit de analyticserver).

as_tenant

Titulaire dont l'installation SPSS Modeler Server est membre (par défaut, il s'agit de ibm).

as_prompt_for_password

Spécifiez N si SPSS Modeler Server est configuré avec le même système d'authentification pour les utilisateurs et les mots de passe que celui utilisé sur Analytic Server ; sinon, spécifiez Y.

Si vous exécutez SPSS Modeler en mode de traitement par lots, ajoutez les arguments `-analytic_server_username {nom_utilisateur_AS}` `-analytic_server_password {mot_de_passe_AS}` à la commande `clemb`.

b. Redémarrez le service SPSS Modeler Server.

Pour pouvoir vous connecter à une installation Analytic Server sur laquelle SSL/TLS est activée, d'autres étapes de configuration de vos installations du serveur et du client SPSS Modeler sont requises.

- a. Accédez à `http{s}://{HOTE}:{PORT}/{RACINE_CONTEXTE}/admin/{TITULAIRE}` et connectez-vous à la console Analytic Server.
- b. Téléchargez le fichier de certification depuis le navigateur et enregistrez-le sur votre système de fichiers.
- c. Ajoutez le fichier de certification à l'environnement d'exécution Java (JRE) de votre installation SPSS Modeler Server et de votre installation SPSS Modeler Client. L'emplacement à mettre à jour est celui sous le sous-répertoire `/jre/lib/security/cacerts` du chemin d'installation SPSS Modeler.

1) Vérifiez que le fichier `cacerts` n'est pas en lecture seule.

2) Utilisez l'outil de clés Modeler livré avec le produit. Cet outil est situé dans le sous-répertoire `/jre/bin/keytool` du chemin d'installation SPSS Modeler.

Exécutez la commande suivante :

```
keytool -import -alias <alias_as> -file <fichier_cert> -keystore "<fichier_cacerts>"
```

Notez que `<alias_as>` est un alias pour le fichier `cacerts`. Vous pouvez utiliser n'importe quel nom dans la mesure où il est unique au fichier `cacerts`.

Un exemple de commande serait similaire à ceci :

```
keytool -import -alias MySSLCertAlias -file C:\Download\as.cer  
-keystore "c:\Program Files\IBM\SPSS\Modeler\{VersionModeler}\jre\lib\security\cacerts"
```

d. Redémarrez SPSS Modeler Server et SPSS Modeler Client .

2. [facultatif] Installez IBM SPSS Modeler - Essentials for R si vous prévoyez d'évaluer les modèles R dans les flux avec sources de données Analytic Server. Vous pouvez télécharger IBM SPSS Modeler - Essentials for R depuis le site <https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=swg-tspssp>.

Chapitre 4. Migration

Analytic Server prend en charge un certain nombre de scénarios de migration de vos paramètres de configuration depuis une installation Analytic Server existante vers une nouvelle installation.

Mise à niveau vers une nouvelle version d'Analytic Server

Si vous disposez d'une installation Analytic Server 1.0.1 existante et avez fait l'acquisition de la version 2.0, vous pouvez migrer vos paramètres de configuration 1.0.1 vers votre installation 2.0. Sous ce scénario, l'outil de migration vous guide tout au long du chemin approprié.

Restriction : Si vous aviez installé la version 1.0, vous devez tout d'abord migrer depuis la version 1.0 vers la version 1.0.1, puis de la version 1.0.1 vers la version 2.0.

Restriction : Vos installations 1.0.1 et 2.0 ne peuvent pas coexister sur le même cluster Hadoop. Si vous configurez votre installation 2.0 afin d'utiliser le même cluster Hadoop que votre installation 1.0.1, cette dernière ne sera plus fonctionnelle.

Création d'une copie de votre installation Analytic Server existante

Si vous disposez d'une installation Analytic Server 2.0 existante, vous pouvez répliquer vos paramètres de configuration dans une nouvelle installation 2.0. Ceci est pratique, par exemple, lorsque vous disposez d'un environnement de développement existant et avez besoin de mettre en place des environnements de test et de production.

Etapes de migration depuis la version 1.0.1 vers la version 2.0

1. Implantez la nouvelle installation d'Analytic Server en suivant les instructions figurant dans «Installation», à la page 6.
2. A la fin de l'installation, choisissez de migrer une configuration. Ceci lancera automatiquement l'outil de migration. Procédez comme suit avant de revenir à l'outil de migration.
 - a. Copiez le répertoire `{RACINE_AS}\tools\configcollector` vers votre nouvelle installation sous l'emplacement correspondant à l'ancienne installation.
 - b. Exécutez l'outil de collecte de la configuration dans votre installation existante en lançant le script **configcollector** situé sous `{RACINE_AS}\tools\configcollector`. Notez l'emplacement du fichier compressé (.zip) résultant.
 - c. Copiez la valeur de la racine d'analyse. Par exemple, si vous aviez installé la 1.0.1 avec comme valeur de racine d'analyse `/user/aeuser/analytic-root` et désirez migrer vers l'emplacement de racine d'analyse `/user/aeuser/analytic-root_new`, exécutez le script de copie suivant.

```
hadoop distcp /user/aeuser/analytic-root /user/aeuser/analytic-root_new
```
3. Vous devrez indiquer à l'outil de migration l'emplacement de la nouvelle racine d'analyse.
4. A la clôture de l'outil de migration, suivez les étapes de post-configuration éventuelles en vous reportant aux instructions de la rubrique Chapitre 3, «Configuration», à la page 9.

Remarque : Si vous aviez configuré R pour son utilisation avec l'installation Analytic Server existante, vous devrez suivre les étapes destinées à le configurer avec la nouvelle installation Analytic Server.

Etapes de migration depuis la version 2.0 vers la version 2.0

1. Exécutez l'outil de collecte de la configuration dans votre installation existante en lançant le script **configcollector** situé sous `{RACINE_AS}\tools\configcollector`. Notez l'emplacement du fichier compressé (.zip) résultant.
2. Implantez la nouvelle installation d'Analytic Server en suivant les instructions figurant dans «Installation», à la page 6.

3. A la fin de l'installation, choisissez de migrer une configuration. Ceci lancera automatiquement l'outil de migration.
4. A la clôture de l'outil de migration, suivez les étapes de post-configuration éventuelles en vous reportant aux instructions de la rubrique Chapitre 3, «Configuration», à la page 9.

Remarque : Si vous aviez configuré R pour son utilisation avec l'installation Analytic Server existante, vous devrez suivre les étapes destinées à le configurer avec la nouvelle installation Analytic Server.

Utilisation de l'outil de migration

L'outil de migration automatise un certain nombre d'étapes de migration de votre configuration existante vers une nouvelle installation, étapes qui devraient autrement être réalisées manuellement. L'outil de migration doit pouvoir accéder à l'environnement HDFS.

Si vous avez utilisé la version de l'interface graphique d'IBM Installation Manager, l'outil de migration se lance automatiquement. Suivez les invites pour configurer votre système.

Vous pouvez également lancer l'outil de migration en exécutant le script **migrationtool** situé sous le répertoire {RACINE_AS}\bin. Par défaut, l'outil détecte s'il doit être lancé dans l'interface graphique ou en mode console. Vous pouvez imposer le mode console en lui transmettant l'argument **-c**.

Bienvenue

Cette étape nécessite de fournir le fichier compressé (zip) compressé contenant la configuration Analytic Server que vous migrez.

Cluster Hadoop

Les étapes de cluster Hadoop configurent Analytic Server pour son utilisation avec votre distribution Hadoop.

Même cluster Hadoop que l'installation d'origine

Il s'agit d'une copie "destructive" dans laquelle la nouvelle installation s'approprie le cluster Hadoop de l'ancienne installation. Cette option est adaptée si vous n'avez plus besoin de l'ancienne installation Analytic Server.

Cluster Hadoop différent

La nouvelle installation est connectée à un cluster Hadoop différent de celui de l'ancienne installation.

Les écrans qui suivent ce choix initial sont similaires à ceux du chemin Standard dans l'outil de configuration.

Hcatalog/Hive

Ces étapes ne sont présentées que si votre installation d'origine comporte des sources de données HCatalog / Hive.

- Dans une migration avec copie vers le même cluster Hadoop, vous avez la possibilité de désactiver Hive.
- Dans une migration avec copie vers un cluster Hadoop différent, vous disposez de la gamme complète d'options du chemin Standard dans l'outil de configuration pour configurer HCatalog / Hive.

JDBC

Dans une migration depuis la version 1.0.1 vers la version 2.0, si vous comptez utiliser des sources de données de base de données, vous devez ajouter à la nouvelle installation des pilotes JDBC via l'outil d'administration.

Dans une migration depuis une version 2.0 vers une version 2.0, les pilotes JDBC éventuels utilisés dans l'installation d'origine sont migrés vers la nouvelle installation.

Serveur d'application

Ces étapes dépendent du type de la nouvelle installation.

- Dans une installation sans haute disponibilité, ces étapes sont ignorées. Vous pouvez configurer une communication sécurisée (SSL) et définir via l'outil d'administration les ports sur lesquels Analytic Server est à l'écoute.
- Dans le cas d'une installation de haute disponibilité, l'outil de migration vous réclame des informations sur votre environnement WebSphere® Application Server Liberty Network Deployment.

Remarque : Si votre installation d'origine était à haute disponibilité, ces étapes sont pré-remplies avec les valeurs de l'installation d'origine.

Zookeeper

L'étape de configuration Zookeeper n'est requise que dans le cadre d'une installation de haute disponibilité. Indiquez la liste des serveurs dans votre ensemble Zookeeper.

Remarque : Si votre installation d'origine était à haute disponibilité, ces étapes sont pré-remplies avec les valeurs de l'installation d'origine.

Référentiel de métadonnées

Ces étapes dépendent du type de la nouvelle installation.

- Dans une installation sans haute disponibilité, vous avez le choix entre Derby et DB2. Si vous choisissez DB2, ce référentiel doit être installé et disponible avant de lancer l'outil de migration. Pour plus d'informations, voir «Prérequis», à la page 5.
- Une configuration d'installation de haute disponibilité doit toujours utiliser DB2 comme référentiel de métadonnées et l'instance DB2 doit être installée et disponible avant le lancement de l'outil de configuration.

Remarque : Si votre installation d'origine était à haute disponibilité, ces étapes sont pré-remplies avec les valeurs de l'installation d'origine.

Analytic Server

L'outil de migration utilise la configuration d'origine pour déterminer le principal faisant office d'administrateur Analytic Server. Ce principal pourra créer, supprimer et gérer les titulaires. Vous pouvez changer de principal via l'outil d'administration.

Récapitulatif

Cette étape vous permet de passer en revue les étapes avant de finaliser la migration.

Post-migration

Cette étape fournit des informations confirmant que la migration a abouti. En cas d'échec, consultez le fichier journal dans le répertoire logs.

Chapitre 5. Apport de modifications à votre système

Utilisation de l'outil d'administration

L'outil d'administration automatise un certain nombre d'étapes de mise à jour de votre configuration qui devraient autrement être réalisées manuellement. Cet outil doit pouvoir accéder à l'environnement HDFS.

Vous pouvez lancer l'outil d'administration en exécutant le script **admintool** situé sous le répertoire `{RACINE_AS}\bin`. Par défaut, l'outil détecte s'il doit être lancé dans l'interface graphique ou en mode console. Vous pouvez imposer le mode console en lui transmettant l'argument **-c**.

Récapitulatif

Le panneau Récapitulatif est toujours affiché avant d'apporter des modifications permanentes au système. Lorsque vous avez terminé une série d'étapes sous une zone Administration, le dernier bouton vous mène au panneau Récapitulatif. Depuis ce panneau, vous pouvez sélectionner une autre zone Administration pour la modifier ou cliquer sur Appliquer pour appliquer les modifications au système.

Cluster Hadoop

Les étapes de cluster Hadoop vous permettent de reconfigurer Analytic Server pour son utilisation avec votre distribution Hadoop.

Remarque : L'outil d'administration vous permet de mettre à jour la version de votre distribution Hadoop si son distributeur prend en charge la mise à jour de la version. Si vous changez de distributeur Hadoop, vous pouvez faire migrer les paramètres de votre installation Analytic Server existante vers une nouvelle installation configurée pour utiliser la nouvelle distribution Hadoop. Pour plus d'informations, voir Chapitre 4, «Migration», à la page 19.

Actualisation simple

Cette procédure est utile si la version de votre distribution Hadoop a changé et que vous avez simplement besoin d'actualiser les fichiers JAR utilisés par Analytic Server.

Actualisation guidée

Cette procédure est utile si d'autres propriétés de votre cluster Hadoop ont changé. Les écrans qui accompagnent l'actualisation guidée sont similaires à ceux de l'outil de configuration.

Hcatalog/Hive

Les étapes Hcatalog/Hive vous permettent de reconfigurer Analytic Server pour son utilisation avec votre installation Hive ou d'activer votre installation Analytic Server existante avec une nouvelle installation Hive.

Si vous disposez d'une installation Hive existante, deux choix sont possibles pour mettre à jour votre installation.

Actualisation simple

Cette procédure est utile si votre version Hive a changé et que vous avez simplement besoin d'actualiser les fichiers JAR utilisés par Analytic Server.

Actualisation guidée

Cette procédure est utile si d'autres propriétés de votre installation Hive ont changé. Les écrans qui accompagnent l'actualisation guidée sont similaires à ceux de l'outil de configuration.

JDBC

Copiez les pilotes des bases de données que vous utilisez vers l'emplacement spécifié et poursuivez les étapes d'administration.

Serveur d'application

Les étapes Serveur d'application dépendent du type d'installation.

- Dans une installation sans haute disponibilité, vous pouvez configurer une communication sécurisée (SSL/TLS) et définir les ports sur lesquels Analytic Server est à l'écoute.
- Dans une installation à haute disponibilité, l'outil d'administration vous permet de noter les modifications éventuelles appliquées à votre environnement WebSphere® Application Server Liberty Network Deployment.

Remarque : Pour ajouter ou supprimer un serveur de quarantaine, vous devez modifier la propriété de configuration `ae.cluster.quarantine.server.name`. Pour plus d'informations, voir «Modification manuelle des propriétés de configuration».

Zookeeper

L'outil d'administration vous permet de modifier la liste des serveurs dans votre ensemble Zookeeper.

Référentiel de métadonnées

Les étapes de référentiel de métadonnées vous permettent de modifier votre instance DB2. Si vous changez de schéma, le nouveau schéma que vous spécifiez doit déjà exister dans la base de données.

Si vous aviez configuré Analytic Server afin d'utiliser Derby comme référentiel de métadonnées et désirez passer à DB2, utilisez l'outil de migration.

Analytic Server

Ces étapes vous permettent de changer le principal faisant office d'administrateur Analytic Server. Ce principal pourra créer, supprimer et gérer les titulaires.

Post-administration

Cette étape fournit des informations confirmant que les modifications de la configuration ont abouti. En cas d'échec, consultez le fichier journal dans le répertoire logs.

Modification manuelle des propriétés de configuration

Certaines propriétés de configuration peuvent être modifiées manuellement.

`config.properties`

Le fichier `config.properties` est situé sous le répertoire `{RACINE_SERVEUR_AS}/configuration`. Vous pouvez modifier ce fichier de deux façons.

- L'outil d'administration, lorsqu'il est lancé en mode interface graphique, vous permet d'apporter des modifications à ce fichier.
- Vous pouvez également modifier manuellement ce fichier.

Dans une installation à haute disponibilité, vous pouvez propager les modifications de `config.properties` via l'outil d'administration. Dans une installation sans haute disponibilité, vous devrez redémarrer Analytic Server.

join.small.data.size

Quantité maximale de données (en octets) qu'Analytic Server tentera de joindre dans un algorithme côté Mappe.

mapred.child.java.opts

Contrôle les tailles de segment de mémoire de JVM pour les mappes et tâches de réduction qui s'exécutent dans Hadoop. Attribuez-lui la valeur la plus grande que les noeuds du cluster sont capables de gérer.

mapred.job.reuse.jvm.num.tasks

Réutilise les machines virtuelles Java (JVM) et tâches de réduction. Augmentez la valeur de cette propriété lorsque le temps de démarrage d'une JVM est élevé. Par exemple, si cette valeur est définie à 1, chaque mappe et chaque tâche de réduction peut s'exécuter dans une nouvelle JVM. Lorsque cette valeur est définie à 5, une nouvelle JVM est créée toutes les 5 tâches.

mmr.threads

Nombre d'unités d'exécution à utiliser pour travaux de réduction de mappe en mémoire (M3R).

Remarque : Les propriétés précédées par le préfixe "mmr" contrôlent la réduction de mappe en mémoire et sont importantes dans les installations utilisant intensivement de petits fichiers. Vous devrez éventuellement tester différentes valeurs afin d'optimiser votre système.

mmr.upper.bound.threshold

Quantité maximale de données qui seront traités par M3R. En cas de volume excédentaire, les données seront traitées par Hadoop.

mmr.synthetic.block.size

Taille de bloc par défaut si le fichier réside sur le système de fichiers local.

mmr.diskspill.threshold

Les données M3R intermédiaires sont consignées sur disque si elles dépassent ce seuil.

mmr.diskspill.output.buffer

Mémoire tampon de sortie de l'associateur. Doit être plus petite que **mmr.diskspill.threshold**.

mmr.taskparallel.targets.threshold

Si le rapport cibles/coeurs est inférieur à ce seuil, les travaux sont traités par M3R.

node.max.jobs

Dans une installation à haute disponibilité, nombre maximal de travaux Analytic Server qu'un membre du cluster peut exécuter simultanément.

ae.cluster.quarantine.server.name

Dans une installation à haute disponibilité, nom du serveur désigné comme serveur de quarantaine.

jvm.options

Le fichier `jvm.options` est situé sous le répertoire `{RACINE_SERVEUR_AS}`. L'ajout ou la modification des lignes suivantes dans `jvm.options` définit la taille de segment mémoire des travaux s'exécutant sur le serveur hébergeant Analytic Server ; c'est-à-dire, pas le serveur Hadoop. Cette option est importante pour l'exécution de petits travaux (M3R) et vous devrez éventuellement tester différentes valeurs afin d'optimiser votre système.

```
-Xms512M  
-Xmx2048M
```

Dans une installation sans haute disponibilité, vous devrez redémarrer Analytic Server. Dans une installation à haute disponibilité, aucune méthode simple ne permet de propager les modifications de `jvm.options` à la collectivité ; Vous devrez mettre à jour manuellement `jvm.options` sur chaque membre du cluster.

Déploiement sur de nouveaux membres du cluster à haute disponibilité

Après avoir rejoint un nouveau membre de serveur vers le cluster dans Liberty, vous pouvez déployer Analytic Server vers les nouveaux membres, par le biais de la commande suivante.

```
{RACINE_AS}/bin/deployRemote.sh -addClusterMembers true
```

Sans l'option **addClusterMembers**, Analytic Server est déployé sur chaque membre du cluster à haute disponibilité.

Chapitre 6. Désinstallation

1. Exécuter IBM Installation Manager (interface graphique ou ligne de commande), sélectionnez cette option pour désinstaller IBM SPSS Analytic Server et suivez les invites. IBM Installation Manager peut également être exécuté en mode silencieux. Pour plus d'informations, reportez-vous à la documentation IBM Installation Manager : http://www-01.ibm.com/support/knowledgecenter/SSDV2W/im_family_welcome.html.

Remarque : Cette opération ne supprime pas les données du dossier Analytic Server sur HDFS. Pour supprimer le dossier de base complet d'Analytic Server, supprimez le répertoire HDFS dans lequel sont enregistrés les artefacts utilisateur d'Analytic Server (entrées et sorties). Il s'agit du répertoire spécifié comme `root.folder.path` dans `{RACINE_SERVEUR_AS}/configuration/config.properties`.

Chapitre 7. Traitement des incidents

Cette section décrit certains problèmes d'installation et de configuration fréquents et explique comment les résoudre.

Clusters à haute disponibilité

java.net.SocketTimeoutException : Expiration du délai d'attente de lecture

Modifiez comme suit la variable d'environnement du délai d'attente de lecture de Liberty ND :

```
export LIBERTYND_READ_TIMEOUT=<millisecondes>
```

où <millisecondes> correspond au nombre de secondes à utiliser pour expiration du délai d'attente de lecture JMX.

java.io.IOException: CWWKX7202E: La valeur du délai d'attente (60 secondes), de la commande ./server start a expiré

Ajoutez ceci au fichier server.xml du serveur contrôleur

```
<!-- Augmentation du délai d'attente de démarrage/d'arrêt
du serveur pour tenir compte des matériels lents -->
<serverCommands startServerTimeout="120" stopServerTimeout="120"/>
```

java.lang.OutOfMemoryError: espace de segment mémoire Java

Ajoutez les lignes suivantes au fichier jvm.options sur chaque membre du cluster à haute disponibilité.

```
-Xms512M
-Xmx2048M
```

Les données de transactions Zookeeper deviennent ingérables

Attribuez au paramètre **autopurge.purgeInterval** dans zoo.cfg la valeur 1 pour permettre des purges automatiques du journal de transactions Zookeeper.

Le service cluster d'analyse perd le contact avec Zookeeper

Examinez et modifiez les paramètres **tickTime**, **initLimit** et **syncLimit** dans zoo.cfg. Par exemple :

```
# Nombre de millisecondes de chaque graduation
tickTime=2000
# Nombre de graduations que la phase de
# synchronisation initiale peut accepter
initLimit=30
# Nombre de graduations pouvant s'écouler entre l'envoi
# d'une demande et son accusé de réception
syncLimit=15
```

Pour plus d'informations, reportez-vous à la documentation Zookeeper : <https://zookeeper.apache.org/doc/r3.3.3/zookeeperAdmin.html>

Les travaux Analytic Server ne reprennent pas

Les travaux Analytic Server ne reprennent pas sous deux situations connues.

1. Lorsqu'un travail Analytic Server échoue en raison de l'échec d'un membre du cluster, le travail est normalement redémarré automatiquement sur un autre membre du cluster. Si le travail ne reprend pas, vérifiez que le cluster de haute disponibilité comprend au moins 4 membres.

2. Lorsque vous mettez au repos un membre de cluster, tous les travaux Analytic Server reprennent normalement sur un autre membre du cluster. Pour garantir la reprise des travaux, spécifiez la valeur `-Dcom.spss.ae.remoteclient.failover.threshold=100` et utilisez le mode distant.

Remarques

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
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.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japon

Le paragraphe suivant ne s'applique ni au Royaume-Uni, ni dans aucun pays dans lequel il serait contraire aux lois locales. LE PRÉSENT DOCUMENT EST LIVRE "EN L'ÉTAT" SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFAÇON AINSI QU'EN CAS DE DEF AUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE. Certaines juridictions n'autorisent pas l'exclusion des garanties implicites, 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 éléments figurant sur ces sites Web ne font pas partie des éléments du présent produit IBM et l'utilisation de ces sites relève de votre seule responsabilité.

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 Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
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 indiquées dans ce document ont été déterminées dans un environnement contrôlé. Par conséquent, les résultats peuvent varier de manière significative selon l'environnement d'exploitation utilisé. Certaines mesures évaluées sur des systèmes en cours de développement ne sont pas garanties sur tous les systèmes disponibles. En outre, elles peuvent résulter d'extrapolations. Les résultats peuvent donc varier. Il incombe aux utilisateurs de ce document de vérifier si ces données sont applicables à leur environnement 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.

Toute instruction relative aux intentions d'IBM pour ses opérations à venir est susceptible d'être modifiée ou annulée sans préavis, et doit être considérée 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. Toute ressemblance avec des noms de personnes, de sociétés ou des données 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 :

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. Toute ressemblance avec des noms de personnes, de sociétés ou des données 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.

© Copyright IBM Corp. _indiquez l'année ou les années_. All rights reserved.

Si vous visualisez ces informations en ligne, il se peut que les photographies et illustrations en couleur n'apparaissent pas à l'écran.

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.

Java ainsi que tous les logos et toutes les marques incluant Java sont des marques d'Oracle et/ou de ses sociétés affiliées.

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.



Imprimé en France