



Informatica®

10.0.0

Informatica® PowerCenter Express

10.0

Guide du réglage de performance

Informatica PowerCenter Express Guide du réglage de performance
10.0.010.0
Novembre 2015

© Copyright Informatica LLC 1998, 2018

Ce logiciel et sa documentation contiennent des informations appartenant à Informatica LLC, protégées par la loi sur le droit d'auteur et fournies dans le cadre d'un accord de licence prévoyant des restrictions d'utilisation et de divulgation. Toute ingénierie inverse du logiciel est interdite. Il est interdit de reproduire ou transmettre sous quelque forme et par quelque moyen que ce soit (électronique, photocopie, enregistrement ou autre) tout ou partie de ce document sans le consentement préalable de Informatica LLC. Ce logiciel peut être protégé par des brevets américains et/ou internationaux, ainsi que par d'autres brevets en attente.

L'utilisation, la duplication ou la divulgation du Logiciel par le gouvernement américain est sujette aux restrictions décrites dans l'accord de licence applicable du logiciel conformément aux documents DFARS 227.7202-1(a) et 227.7702-3(a) (1995), DFARS 252.227-7013(1)(ii) (OCT 1988), FAR 12.212(a) (1995), FAR 52.227-19 ou FAR 52.227-14 (ALT III) le cas échéant.

Les informations dans ce produit ou cette documentation sont sujettes à modification sans préavis. Si vous rencontrez des problèmes dans ce produit ou la documentation, veuillez nous en informer par écrit.

Informatica, Informatica Platform, Informatica Data Services, PowerCenter, PowerCenterRT, PowerCenter Connect, PowerCenter Data Analyzer, PowerExchange, PowerMart, Metadata Manager, Informatica Data Quality, Informatica Data Explorer, Informatica B2B Data Transformation, Informatica B2B Data Exchange Informatica On Demand, Informatica Identity Resolution, Informatica Application Information Lifecycle Management, Informatica Complex Event Processing, Ultra Messaging et Informatica Master Data Management sont des marques de commerce ou des marques déposées d'Informatica LLC aux États-Unis et dans d'autres juridictions du monde. Tous les autres noms de société ou de produit peuvent être des marques de commerce ou des marques déposées de leurs détenteurs respectifs.

Des portions de ce logiciel et/ou de la documentation sont sujettes au copyright détenu par des tierces parties, dont Copyright DataDirect Technologies. Tous droits réservés. Copyright © Sun Microsystems. Tous droits réservés. Copyright © RSA Security Inc. Tous droits réservés. Copyright © Ordinal Technology Corp. Tous droits réservés. Copyright © Aandacht c.v. Tous droits réservés. Copyright Genivia, Inc. Tous droits réservés. Copyright Isomorphic Software. Tous droits réservés. Copyright © Meta Integration Technology, Inc. Tous droits réservés. Copyright © Intalio. Tous droits réservés. Copyright © Oracle. Tous droits réservés. Copyright © Adobe Systems Incorporated. Tous droits réservés. Copyright © DataArt, Inc. Tous droits réservés. Copyright © ComponentSource. Tous droits réservés. Copyright © Microsoft Corporation. Tous droits réservés. Copyright © Rogue Wave Software, Inc. Tous droits réservés. Copyright © Teradata Corporation. Tous droits réservés. Copyright © Yahoo! Inc. Tous droits réservés. Copyright © Glyph & Cog, LLC. Tous droits réservés. Copyright © Thinkmap, Inc. Tous droits réservés. Copyright © Clearpace Software Limited. Tous droits réservés. Copyright © Information Builders, Inc. Tous droits réservés. Copyright © OSS Nokalva, Inc. Tous droits réservés. Copyright Edifecs, Inc. Tous droits réservés. Copyright Cleo Communications, Inc. Tous droits réservés. Copyright © International Organization for Standardization 1986. Tous droits réservés. Copyright © ej-technologies GmbH. Tous droits réservés. Copyright © Jaspersoft Corporation. Tous droits réservés. Copyright © International Business Machines Corporation. Tous droits réservés. Copyright © yWorks GmbH. Tous droits réservés. Copyright © Lucent Technologies. Tous droits réservés. Copyright © Université de Toronto. Tous droits réservés. Copyright © Daniel Veillard. Tous droits réservés. Copyright © Unicode, Inc. Copyright IBM Corp. Tous droits réservés. Copyright © MicroQuill Software Publishing, Inc. Tous droits réservés. Copyright © PassMark Software Pty Ltd. Tous droits réservés. Copyright © LogiXML, Inc. Tous droits réservés. Copyright © 2003-2010 Lorenzi Davide. Tous droits réservés. Copyright © Red Hat, Inc. Tous droits réservés. Copyright © The Board of Trustees of the Leland Stanford Junior University. Tous droits réservés. Copyright © EMC Corporation. Tous droits réservés. Copyright © Flexera Software. Tous droits réservés. Copyright © Jinfonet Software. Tous droits réservés. Copyright © Apple Inc. Tous droits réservés. Copyright © Telerix Inc. Tous droits réservés. Copyright © BEA Systems. Tous droits réservés. Copyright © PDFlib GmbH. Tous droits réservés. Copyright © Orientation in Objects GmbH. Tous droits réservés. Copyright © Tanuki Software, Ltd. Tous droits réservés. Copyright © Ricebridge. Tous droits réservés. Copyright © Sencha, Inc. Tous droits réservés. Copyright © Scalable Systems, Inc. Tous droits réservés. Copyright © jQWidgets. Tous droits réservés. Copyright © Tableau Software, Inc. Tous droits réservés. Copyright © MaxMind, Inc. Tous droits réservés. Copyright © TMate Software s.r.o. Tous droits réservés. Copyright © MapR Technologies Inc. Tous droits réservés. Copyright © Amazon Corporate LLC. Tous droits réservés. Copyright © Highsoft. Tous droits réservés. Copyright © Python Software Foundation. Tous droits réservés. Copyright © BeOpen.com. Tous droits réservés. Copyright © CNRI. Tous droits réservés.

Ce produit inclut des logiciels développés par Apache Software Foundation (<http://www.apache.org/>), et/ou d'autres logiciels sous licence et sous diverses versions Apache License (la « Licence »). Vous pouvez obtenir une copie de ces licences à l'adresse suivante : <http://www.apache.org/licenses/>. Sauf dispositions contraires de la loi en vigueur ou accord écrit, le logiciel distribué sous cette licence est livré « EN L'ÉTAT », SANS GARANTIE NI CONDITION D'AUCUNE SORTIE, expresse ou implicite. Se reporter aux Licences pour la langue spécifique régissant les droits et limitations dans le cadre des Licences.

Ce produit inclut des logiciels développés par Mozilla (<http://www.mozilla.org/>), copyright de logiciel The JBoss Group, LLC, tous droits réservés ; copyright de logiciel © 1999-2006 de Bruno Lowagie et Paulo Soares et d'autres logiciels sous licence et sous diverses versions du GNU Lesser General Public License Agreement, accessible sur <http://www.gnu.org/licenses/lgpl.html>. Les matériaux sont fournis gratuitement par Informatica, « en l'état », sans garantie d'aucune sorte, expresse ou implicite, notamment les garanties implicites de conformité légale et d'usage normal.

Le produit inclut les logiciels ACE(TM) et TAO(TM), copyright Douglas C. Schmidt et son groupe de recherche à Washington University, University of California, Irvine, et Vanderbilt University, Copyright (©) 1993-2006, tous droits réservés.

Ce produit inclut des logiciels développés par OpenSSL Project pour une utilisation dans OpenSSL Toolkit (copyright The OpenSSL Project. Tous droits réservés) et la redistribution de ce logiciel est sujette aux termes publiés sur <http://www.openssl.org> et <http://www.openssl.org/source/license.html>.

Ce produit inclut le logiciel Curl, copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>. Tous Droits Réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://curl.haxx.se/docs/copyright.html>. L'autorisation d'utiliser, copier, modifier et distribuer ce logiciel à toute fin, avec ou sans rémunération, est accordée par les présentes, à la condition que la notification de copyright ci-dessus et cette notification d'autorisation apparaissent dans toutes les copies.

Le produit inclut des logiciels sous copyright 2001-2005 (©) MetaStuff, Ltd. Tous droits réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://www.dom4j.org/license.html>.

Le produit inclut des logiciels sous copyright © 2004-2007, The Dojo Foundation. Tous Droits Réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://dojotoolkit.org/license>.

Ce produit inclut le logiciel ICU sous copyright de International Business Machines Corporation et autres. Tous Droits Réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://source.icu-project.org/repos/icu/icu/trunk/license.html>.

Ce produit inclut des logiciels sous copyright © 1996-2006 Per Bothner. Tous Droits Réservés. Votre droit à utiliser de tels matériels est défini dans la licence qui peut être consultée sur <http://www.gnu.org/software/kawa/Software-License.html>.

Ce produit inclut le logiciel OSSP UUID sous copyright © 2002 Ralf S. Engelschall, copyright © 2002 The OSSP Project Copyright © 2002 Cable & Wireless Deutschland. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://www.opensource.org/licenses/mit-license.php>.

Ce produit inclut des logiciels développés par Boost (<http://www.boost.org/>) ou sous licence de logiciel Boost. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur http://www.boost.org/LICENSE_1_0.txt.

Ce produit inclut des logiciels sous copyright © 1997-2007 University of Cambridge. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://www.pcre.org/license.txt>.

Ce produit inclut des logiciels sous copyright © 2007 The Eclipse Foundation. Tous Droits Réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://www.eclipse.org/org/documents/epl-v10.php> et <http://www.eclipse.org/org/documents/edl-v10.php>.

Ce produit comprend des logiciels sous licence dont les conditions se trouvent aux adresses : <http://www.tcl.tk/software/tcltk/license.html>, <http://www.bosrup.com/web/overlib/?License>, <http://www.stlport.org/doc/license.html>, <http://asm.ow2.org/license.html>, <http://www.cryptix.org/LICENSE.TXT>, <http://hsqldb.org/web/hsqldbLicense.html>, <http://httpunit.sourceforge.net/doc/license.html>, <http://jung.sourceforge.net/license.txt>, http://www.gzip.org/zlib/zlib_license.html, <http://www.openldap.org/software/release/license.html>, <http://www.libssh2.org>, <http://slf4j.org/license.html>, <http://www.sente.ch/software/OpenSourceLicense.html>, <http://fusesource.com/downloads/license-agreements/fuse-message-broker-v-5-3-license-agreement>; <http://antlr.org/license.html>; <http://aopalliance.sourceforge.net/>; <http://www.bouncycastle.org/licence.html>; <http://www.jgraph.com/jgraphdownload.html>; <http://www.jcraft.com/jsch/LICENSE.txt>; http://jotm.objectweb.org/bsd_license.html; <http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231>; <http://www.slf4j.org/license.html>; <http://nanoxml.sourceforge.net/orig/copyright.html>; <http://www.json.org/license.html>; <http://forge.ow2.org/projects/javaservice/>, <http://www.postgresql.org/about/licence.html>, <http://www.sqlite.org/copyright.html>, <http://www.tcl.tk/software/tcltk/license.html>, <http://www.jaxen.org/faq.html>, <http://www.jdom.org/docs/faq.html>, <http://www.slf4j.org/license.html>, <http://www.iodbc.org/dataspace/iodbc/wiki/IODBC/License>; <http://www.keplerproject.org/md5/license.html>; <http://www.toedter.com/en/jcalendar/license.html>; <http://www.edankert.com/bounce/index.html>; <http://www.net-snmp.org/about/license.html>; <http://www.openmdx.org/#FAQ>; http://www.php.net/license/3_01.txt; <http://srp.stanford.edu/license.txt>; <http://www.schneier.com/blowfish.html>; <http://www.jmock.org/license.html>; <http://xsom.java.net>; <http://benalman.com/about/license/>; <https://github.com/CreateJS/EaselJS/blob/master/src/easeljs/display/Bitmap.js>; <http://www.h2database.com/html/license.html#summary>; <http://jsoncpp.sourceforge.net/LICENSE>; <http://jdbc.postgresql.org/license.html>; <http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto>; <https://github.com/rantav/hector/blob/master/LICENSE>; <http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html>; <http://jibx.sourceforge.net/jibx-license.html>; <https://github.com/lyokato/libgeohash/blob/master/LICENSE>; <https://github.com/hjiang/jsonxx/blob/master/LICENSE>; <https://code.google.com/p/lz4/>; <https://github.com/jedisct1/libsodium/blob/master/LICENSE>; <http://one-jar.sourceforge.net/index.php?page=documents&file=license>; <https://github.com/EsotericSoftware/kryo/blob/master/license.txt>; <http://www.scala-lang.org/license.html>; <https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt>; <http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html>; <https://aws.amazon.com/asl/>; <https://github.com/twbs/bootstrap/blob/master/LICENSE> et <https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt>.

Ce produit inclut un logiciel sous licence Academic Free License (<http://www.opensource.org/licenses/afl-3.0.php>), licence Common Development Distribution License (<http://www.opensource.org/licenses/cddl1.php>) licence Common Public License (<http://www.opensource.org/licenses/cpl1.0.php>), licence Sun Binary Code License Agreement Supplemental License Terms, licence BSD (<http://www.opensource.org/licenses/bsd-license.php>), le nouvelle licence BSD License (<http://opensource.org/licenses/BSD-3-Clause>), la licence MIT (<http://www.opensource.org/licenses/mit-license.php>), la licence Artistic License (<http://www.opensource.org/licenses/artistic-license-1.0>) et la licence publique du développeur initial Version 1.0 (<http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/>).

Ce produit inclut des logiciels sous copyright © 2003-2006 Joe Walnes, 2006-2007 XStream Committers. Tous Droits Réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://xstream.codehaus.org/license.html>. Ce produit inclut des logiciels développés par Indiana University Extreme! Lab. Pour plus d'informations, veuillez vous rendre sur <http://www.extreme.indiana.edu/>.

Ce produit inclut des logiciels sous copyright © 2013 Frank Balluffi et Markus Moeller. Tous droits réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions de la licence MIT.

Consultez les brevets applicables à l'adresse <https://www.informatica.com/legal/patents.html>.

EXCLUSION DE RESPONSABILITÉ : Informatica LLC fournit cette documentation « en l'état », sans garantie d'aucune sorte, explicite ou implicite, notamment les garanties implicites de non-infraction, de conformité légale ou d'usage normal. Informatica LLC ne garantit pas que ce logiciel et cette documentation sont exempts d'erreurs. Les informations fournies dans ce logiciel ou cette documentation peuvent inclure des inexactitudes techniques ou des erreurs typographiques. Les informations contenues dans ce logiciel et sa documentation sont sujettes à modification à tout moment sans préavis.

AVIS

Ce produit Informatica (le « Logiciel ») inclut certains pilotes (les « Pilotes DataDirect ») de DataDirect Technologies, une société de Progress Software Corporation (« DataDirect ») qui sont sujets aux conditions suivantes :

1. LES PILOTES DATADIRECT SONT FOURNIS « EN L'ÉTAT », SANS GARANTIE D'AUCUNE SORTE, EXPRESSE OU IMPLICITE, NOTAMMENT LES GARANTIES IMPLICITES DE CONFORMITÉ LÉGALE, D'USAGE NORMAL ET DE NON-INFRACTION.
2. DATADIRECT OU SES FOURNISSEURS TIERS NE POURRONT EN AUCUN CAS ÊTRE TENUS RESPONSABLES ENVERS LE CLIENT UTILISATEUR FINAL DE TOUT DOMMAGE DIRECT, ACCESSOIRE, INDIRECT, SPÉCIAL, CONSÉCUTIF OU AUTRE RÉSULTANT DE L'UTILISATION DES PILOTES ODBC, QU'ILS SOIENT INFORMÉS OU NON À L'AVANCE DE LA POSSIBILITÉ DE TELS DOMMAGES. CES LIMITATIONS S'APPLIQUENT À TOUTES LES CAUSES D'ACTION, NOTAMMENT TOUTE INFRACTION AU CONTRAT, INFRACTION À LA GARANTIE, NÉGLIGENCE, RESPONSABILITÉ STRICTE, REPRÉSENTATION INCORRECTE ET AUTRES TORTS.

Date de publication: 2018-07-03

Sommaire

Préface.....	8
Ressources Informatica.	8
Portail Mon support Informatica.	8
Documentation Informatica.	9
Matrices de disponibilité de produit Informatica.	9
Site Web Informatica.	9
Bibliothèque de procédures Informatica.	9
Base de connaissances Informatica.	9
Canal YouTube du support Informatica.	9
Informatica Marketplace.	10
Informatica Velocity.	10
Support client international Informatica.	10
Chapitre 1: Présentation du réglage des performances.....	11
Présentation du réglage de performance.	11
Goulots d'étranglement de cible.	12
Goulots d'étranglement de source.	12
Goulots d'étranglement de mappage.	13
Goulots d'étranglement du système informatique.	13
Identification des goulots d'étranglement sous Windows.	13
Identification des goulots d'étranglement sous UNIX.	14
Goulots d'étranglement d'exécution.	14
Goulots d'étranglement de l'optimisation du service de données SQL.	14
Goulots d'étranglement de l'optimisation du service Web.	15
Goulots d'étranglement de la connexion.	15
Chapitre 2: Optimisation de cible.....	16
Présentation de l'optimisation de cible.	16
Optimisation de cible de fichier plat.	16
Intervalles de point de contrôle de base de données.	17
Chargements en bloc.	17
Optimisation de cible de base de données.	18
Chapitre 3: Optimisation de source.....	19
Présentation de l'optimisation de source.	19
Optimisation de source de fichier plat.	20
Optimisation de requête.	20
Filtres conditionnels.	21
Sélectionner distinct.	21
Indicateurs.	22

Règles et instructions concernant les indicateurs.	22
Création d'indicateurs.	23
Contraintes.	23
Configuration de contraintes.	24
Optimisation d'objet de données personnalisé.	24
Optimisation de source de base de données.	25
Chapitre 4: Optimisation de transformation.	26
Optimisation de transformation.	26
Optimisation de la transformation Agrégation.	27
Transformation Processeur de données.	27
Optimisation d'expression.	28
Optimisation de la transformation Java.	30
Optimisation de sélection précoce avec la transformation Java.	30
Optimisation push-into avec la transformation Java.	31
Optimisation de la transformation Jointure.	32
Optimisation de la transformation Recherche.	33
Optimisation de la transformation Trieur.	36
Optimisation de la transformation SQL.	36
Optimisation de sélection précoce avec la transformation SQL.	37
Optimisation push-into avec la transformation SQL.	37
Cache de transformation.	38
Élimination d'erreurs de transformation.	39
Effets secondaires de transformation.	39
Optimisation de la transformation Consommateur de service Web.	40
Optimisation de sélection précoce avec la transformation Consommateur de service Web.	41
Optimisation push-into avec la transformation Consommateur de service Web.	41
Chapitre 5: Optimisation du mappage.	44
Présentation de l'optimisation du mappage.	44
Méthodes d'optimisation.	45
Niveaux d'optimisation.	45
Optimisations de filtre.	46
Méthode d'optimisation de projection précoce.	46
Méthode d'optimisation de prédicat.	47
Méthode d'optimisation basée sur le coût.	48
Méthode d'optimisation de jointure dataship.	49
Méthode d'optimisation de semi-jointure.	50
Méthode d'optimisation de sélection précoce.	51
Méthode d'optimisation globale des prédicats.	51
Méthode d'optimisation de nettoyage de branche.	51
Méthode d'optimisation push-into.	52
Optimisation du refoulement.	52

Optimisation du refoulement complet.	53
Refoulement de la source.	54
Règles et instructions d'optimisation du refoulement.	54
Lecture à un seul passage.	54
Optimisation de filtre.	55
Optimisation de conversion du type de données.	56
Suivi d'erreur.	56
Chapitre 6: Optimisation du mappage partitionné.	58
Présentation de l'optimisation du mappage partitionné.	58
Utiliser plusieurs processeurs.	59
Augmenter la valeur de parallélisme maximal.	59
Optimiser les fichiers plats pour le partitionnement.	60
Optimiser les sources de fichiers plats pour le partitionnement.	60
Optimiser les cibles de fichiers plats pour le partitionnement.	60
Optimiser les bases de données relationnelles pour le partitionnement.	61
Optimiser la base de données source pour le partitionnement.	61
Optimiser la base de données cible pour le partitionnement.	61
Optimiser les transformations pour le partitionnement.	62
Chapitre 7: Optimisation du temps d'exécution.	64
Présentation de l'optimisation de l'exécution.	64
Optimisation du service d'application.	64
Optimisation du service Analyst.	65
Optimisation du service d'intégration de données.	65
Optimisation du Model Repository Service.	66
Statistiques de surveillance.	66
Allocation de mémoire.	68
Mise en cache de l'objet de données.	69
Types de données pour les tables de mise en cache.	70
Optimisation de cache d'objet de données.	72
Optimisation du système.	73
Chapitre 8: Optimisation du service de données SQL.	74
Présentation de l'optimisation du service de données SQL.	74
Optimisation d'outil client tiers.	75
Niveaux d'optimisation du service de données SQL.	75
Configuration du niveau d'optimisation du service de données SQL pour la prévisualisation des données.	76
Configuration du niveau d'optimisation pour les services de données SQL déployés.	76
Plan de requête du service de données SQL.	77
Consultation d'un plan de requête SQL.	78
Propriétés du service de données SQL pour la mémoire et les demandes simultanées.	79

Cache de l'ensemble de résultats pour un service de données SQL.	80
Propriétés de cache de l'ensemble de résultats de service de données SQL.	81
Activation de la mise en cache de l'ensemble des résultats pour un service de données SQL.	81
Conservation des données virtuelles dans les tables temporaires.	82
Implémentation des tables temporaires.	82
Chapitre 9: Optimisation du service Web.	83
Présentation de l'optimisation du service Web.	83
Optimisation des requêtes HTTP.	84
Compression de messages du service Web.	84
Niveau d'optimisation du service Web	84
Configuration du niveau d'optimisation du service Web pour la prévisualisation des données	85
Configuration du niveau d'optimisation pour les services Web déployés.	85
Propriétés des services Web pour la mémoire et les demandes simultanées	86
Exemple de configuration du service d'intégration de données pour des demandes de service Web simultanées	88
Propriété du service Web pour configurer une instance DTM active.	88
Mise en cache de l'ensemble de résultats du service Web.	89
Activation de la mise en cache de l'ensemble des résultats pour un service Web.	89
Gestion du journal du service Web.	90
Chapitre 10: Optimisation des connexions.	91
Présentation de l'optimisation de connexions.	91
Pooling de connexions.	91
Propriétés de pooling dans les objets de connexion.	92
Taille du paquet réseau de base de données.	93
Index.	94

Préface

Le *Guide du réglage de performance Informatica* est destiné aux administrateurs et aux développeurs qui souhaitent améliorer les performances. Ce guide suppose que vous connaissez les systèmes d'exploitation, les réseaux, les outils client, les concepts de base de données relationnelle et les fichiers plats de votre environnement. Pour de plus amples informations sur le réglage de performance de base de données non couvert dans ce guide, consultez la documentation de vos produits de base de données.

Le *Guide du réglage de performance PowerCenter Express* est destiné aux administrateurs et aux développeurs qui souhaitent améliorer les performances. Ce guide suppose que vous connaissez les systèmes d'exploitation, les réseaux, les outils client, les bases de données relationnelles et les fichiers plats de votre environnement. Pour de plus amples informations sur le réglage de performance de base de données non couvert dans ce guide, consultez la documentation de vos produits de base de données.

Ressources Informatica

Portail Mon support Informatica

En tant que client Informatica, la première étape pour nous atteindre est via le portail Mon support Informatica sur le site <https://mysupport.informatica.com>. Le portail Mon support est la plus grande plateforme de collaboration d'intégration de données en ligne avec plus de 100 000 clients et partenaires d'Informatica à travers le monde.

En tant que membre, vous pouvez :

- Accéder à toutes les ressources Informatica d'un emplacement.
- Vérifier votre cas de support.
- Faire des recherches dans la base de connaissances, trouver de la documentation sur les produits, accéder aux documents de procédure et regarder des vidéos de support.
- Rechercher votre réseau de groupe d'utilisateurs local Informatica et collaborer avec vos pairs.

En tant que membre, vous pouvez :

- Accéder à toutes les ressources Informatica d'un emplacement.
- Faire des recherches dans la base de connaissances, trouver de la documentation sur les produits, accéder aux documents de procédure et regarder des vidéos de support.
- Rechercher votre réseau de groupe d'utilisateurs local Informatica et collaborer avec vos pairs.

Documentation Informatica

L'équipe Documentation d'Informatica s'efforce de fournir une documentation précise et utilisable. N'hésitez pas à contacter l'équipe Documentation d'Informatica par courriel à l'adresse infa_documentation@informatica.com pour lui faire part de vos questions, commentaires ou suggestions concernant cette documentation. Ces commentaires et suggestions nous permettront d'améliorer notre documentation. Veuillez préciser si vous acceptez d'être contacté au sujet de ces commentaires.

L'équipe Documentation met à jour la documentation chaque fois que nécessaire. Pour obtenir la toute dernière version de la documentation concernant votre produit, consultez la Documentation de produit sur <https://mysupport.informatica.com>.

Matrices de disponibilité de produit Informatica

Les matrices de disponibilité de produit (PAM) indiquent les versions des systèmes d'exploitation, les bases de données et les autres types de sources et cibles de données pris en charge par une version d'un produit. Vous pouvez consulter les PAM sur le portail Mon Support Informatica à l'adresse <https://mysupport.informatica.com>.

Site Web Informatica

Vous pouvez accéder au site Web d'entreprise Informatica sur <https://www.informatica.com>. Le site contient des informations sur Informatica, son expertise, les événements à venir et les bureaux de vente. Vous y trouverez aussi des informations sur ses produits et ses partenaires. Les rubriques de service du site fournissent des informations importantes sur le support technique, la formation et l'éducation, ainsi que les services d'implémentation.

Bibliothèque de procédures Informatica

En tant que client Informatica, vous avez accès à la bibliothèque de procédures Informatica sur <https://mysupport.informatica.com>. La bibliothèque de procédures Informatica est une collection de ressources destinée à vous familiariser avec les produits Informatica et leurs fonctionnalités. Elle regroupe des articles et des démonstrations interactives qui permettent de résoudre des problèmes courants et de comparer les fonctionnalités et les comportements, et qui vous guident lors de la réalisation de tâches concrètes spécifiques.

Base de connaissances Informatica

En tant que client Informatica, vous avez accès à la base de connaissances Informatica sur <https://mysupport.informatica.com>. Utilisez la base de connaissances pour rechercher des solutions documentées aux problèmes techniques connus concernant les produits Informatica. Vous y trouverez également la réponse aux questions les plus fréquentes, des livres blancs et des conseils techniques. N'hésitez pas à contacter l'équipe Base de connaissances Informatica par courriel à l'adresse KB_Feedback@informatica.com pour lui faire part de vos questions, commentaires et suggestions concernant la base de connaissances.

Canal YouTube du support Informatica

Vous pouvez accéder au canal YouTube du support Informatica sur <http://www.youtube.com/user/INFASupport>. Le canal YouTube du support Informatica contient des vidéos concernant les solutions qui vous guident dans l'exécution de tâches spécifiques. Si vous avez des questions, commentaires ou suggestions concernant le canal YouTube du support Informatica, contactez

l'équipe de support YouTube par courriel à l'adresse supportvideos@informatica.com ou envoyez un tweet à @INFASupport.

Informatica Marketplace

Informatica Marketplace est un forum où développeurs et partenaires peuvent partager des solutions qui permettent d'augmenter, d'étendre ou d'améliorer les implémentations d'intégration de données. En tirant profit des centaines de solutions disponibles sur Marketplace, vous pouvez améliorer votre productivité et accélérer le temps d'implémentation de vos projets. Vous pouvez accéder à Informatica Marketplace à l'adresse <http://www.informaticamarketplace.com>.

Informatica Velocity

Vous pouvez accéder à Informatica Velocity à l'adresse <https://mysupport.informatica.com>. Développé à partir de l'expérience concrète de centaines de projets de gestion de données, Informatica Velocity représente le savoir collectif de nos consultants, qui ont travaillé avec des entreprises du monde entier pour planifier, développer, déployer et tenir à jour des solutions de gestion des données efficaces. Si vous avez des questions, des commentaires et des suggestions sur Informatica Velocity, contactez le support des services professionnels Informatica à l'adresse ips@informatica.com.

Support client international Informatica

Vous pouvez contacter un centre de support client par téléphone ou via l'assistance en ligne.

L'assistance en ligne requiert un nom d'utilisateur et un mot de passe. Vous pouvez demander un nom d'utilisateur et un mot de passe sur <http://mysupport.informatica.com>.

Les numéros de téléphone du support client international Informatica sont disponibles sur le site Web Informatica à l'adresse <http://www.informatica.com/us/services-and-training/support-services/global-support-centers/>.

CHAPITRE 1

Présentation du réglage des performances

Ce chapitre comprend les rubriques suivantes :

- [Présentation du réglage de performance, 11](#)
- [Goulots d'étranglement de cible, 12](#)
- [Goulots d'étranglement de source, 12](#)
- [Goulots d'étranglement de mappage, 13](#)
- [Goulots d'étranglement du système informatique, 13](#)
- [Goulots d'étranglement d'exécution, 14](#)
- [Goulots d'étranglement de l'optimisation du service de données SQL, 14](#)
- [Goulots d'étranglement de l'optimisation du service Web, 15](#)
- [Goulots d'étranglement de la connexion, 15](#)

Présentation du réglage de performance

L'objectif du réglage de performance est d'éliminer les goulots d'étranglement des performances. Un goulot d'étranglement est une zone du mappage qui fonctionne le plus fréquemment et a le débit le plus faible. Un goulot d'étranglement réduit les performances globales du mappage.

Pour optimiser un mappage, identifiez un goulot d'étranglement de performances, éliminez-le, puis identifiez le suivant. Optimisez un composant de mappage à la fois. Vous pouvez chronométrer un mappage avant et après la modification pour vérifier que l'optimisation a un impact sur les performances.

Vous pouvez optimiser les composants de mappage dans l'ordre suivant :

1. Cibles
2. Sources
3. Mappages
4. Transformations
5. Environnement Informatica dans l'outil Administrator
6. Système informatique
7. Service de données ou service Web

Utilisez les méthodes suivantes pour identifier des goulots d'étranglement de performances :

- Exécutez des mappages test. Vous pouvez configurer un mappage test pour lire une source du fichier plat ou écrire dans une cible du fichier plat et identifier les goulots d'étranglement dans la source et la cible.
- Analysez les détails des performances. Analysez les détails des performances, tels que les méthodes d'optimisation, pour déterminer où les performances de mappage diminuent.
- Surveillez les performances du système. Vous pouvez utiliser des outils de surveillance système pour afficher le pourcentage d'utilisation du processeur, d'attentes d'E/S, de pagination et d'utilisation des ressources système.

Goulots d'étranglement de cible

Les goulots d'étranglement de cible sont des baisses de performance lorsque le Data Integration Service écrit dans une cible. Les goulots d'étranglement de cible peuvent intervenir lorsque la base de données utilise des petits intervalles de point de contrôle ou des petites tailles de paquet réseau de base de données.

Le goulot d'étranglement de performance le plus courant survient lorsque le Data Integration Service écrit dans une base de données cible. Si la base de données utilise des petits intervalles de point de contrôle, le traitement de la base de données ralentit plus souvent pour écrire un point de contrôle. Des tailles de paquet réseau de base de données peu volumineuses peuvent entraîner des goulots d'étranglement. Vous pouvez permettre à de plus grands paquets de données de traverser le réseau simultanément.

Pour identifier un goulot d'étranglement de cible, vous pouvez créer une copie du mappage avec une cible de fichier plat plutôt qu'une cible de base de données. Si les performances augmentent considérablement, vous avez un goulot d'étranglement de cible. Si le mappage écrit déjà dans une cible de fichier plat, vous n'avez probablement pas de goulot d'étranglement de cible.

Goulots d'étranglement de source

Les goulots d'étranglement de source sont des baisses de performances lorsque le Data Integration Service lit à partir d'une base de données source. Les goulots d'étranglement de source peuvent survenir lorsque la requête source n'est pas efficace ou lorsque les tailles de paquet réseau de base de données sont petites.

Lorsque le mappage lit à partir d'une source relationnelle, vous pouvez utiliser les méthodes suivantes pour identifier les goulots d'étranglement de source :

- Ajoutez une transformation Filtre au mappage. Ajoutez une transformation Filtre après la source. Définissez la condition de filtre sur faux pour que la transformation Filtre ne renvoie aucune donnée. Si le mappage prend à peu près le même temps, le mappage a un goulot d'étranglement de source.
- Créez un mappage de test de lecture. Faites une copie du mappage mais retirez toutes les transformations, jointures ou requêtes. Connectez la source à une cible. Si les performances du mappage sont similaires à celles du mappage d'origine, vous avez un goulot d'étranglement de source.
- Exécutez la requête de lecture directement sur la base de données source. Copiez la requête de lecture à partir du journal de mappage. Exécutez à nouveau la requête sur la base de données source avec un outil de requête tel que isql. Chronométrez le temps d'exécution et le temps mis par la requête pour renvoyer une ligne.

Goulots d'étranglement de mappage

Si vous déterminez que vous n'avez pas de goulot d'étranglement de source ou de cible, vous pouvez avoir un goulot d'étranglement de mappage. Une petite taille de cache, une mémoire tampon faible et des petits intervalles de validation peuvent entraîner des goulots d'étranglement de mappage.

Pour identifier un goulot d'étranglement de mappage, analysez les détails des performances dans le journal de mappage. Les détails des performances incluent des informations sur chaque transformation, telles que le nombre de lignes d'entrée, de lignes de sortie et de lignes d'erreur.

Vous pouvez également ajouter une transformation Filtre avant chaque définition de cible. Définissez la condition de filtre sur faux pour que la transformation filtre ne charge pas de données dans les tables cible. Si l'exécution du nouveau mappage prend le même temps que le mappage d'origine, il y a un goulot d'étranglement de mappage.

Goulots d'étranglement du système informatique

Vous pouvez afficher l'utilisation des ressources lorsque vous exécutez des services Informatica sous Windows ou UNIX. Sous Windows, utilisez le Gestionnaire des tâches. UNIX comporte plusieurs outils permettant de consulter les performances.

Identification des goulots d'étranglement sous Windows

Vous pouvez afficher les onglets Performances et Processus du Gestionnaire des tâches pour consulter les informations système. L'onglet Performances du Gestionnaire des tâches fournit un aperçu de l'utilisation du processeur et de la quantité totale de mémoire utilisée. Utilisez l'Analyseur de performances pour afficher des informations plus détaillées.

Le tableau suivant décrit les informations système que vous pouvez utiliser dans l'Analyseur de performances Windows pour créer un graphique :

Propriété	Description
Pourcentage de temps processeur	Si vous avez plusieurs processeurs, surveille le pourcentage de temps processeur de chaque processeur.
Pages/seconde	Si la valeur pages/seconde est supérieure à cinq, la sollicitation de la mémoire est peut-être trop importante, ce que l'on appelle thrashing.
Pourcentage de temps des disques physiques	Pourcentage de temps pendant lequel le disque physique est occupé à effectuer des requêtes de lecture ou d'écriture.
Longueur de la file d'attente des disques physiques	Nombre d'utilisateurs qui attendent d'accéder au même disque.
Nombre total d'octets par seconde du serveur	Le serveur a envoyé des données au réseau et reçu des données du réseau.

Identification des goulots d'étranglement sous UNIX

Utilisez les outils suivants pour identifier les goulots d'étranglement sous UNIX :

- **top.** Affiche la performance globale du système. Cet outil affiche l'utilisation du processeur, l'utilisation de la mémoire et l'utilisation de l'espace d'échange pour le système et pour les processus individuels exécutés sur le système.
- **iostat.** Surveille l'opération de chargement pour chaque disque connecté au serveur de base de données. Iostat affiche le pourcentage de temps pendant lequel le disque est physiquement actif. Si vous utilisez des baies de disques, utilisez les utilitaires fournis avec les baies de disques plutôt qu'iostat.
- **vmstat.** Surveille les actions d'échange de disque.
- **sar.** Affiche des rapports d'activité du système détaillés sur le processeur, la mémoire et l'utilisation du disque. Vous pouvez utiliser cet outil pour surveiller le chargement du processeur. Il fournit un pourcentage d'utilisation sur les utilisateurs, le système, la durée d'inactivité et le temps d'attente. Vous pouvez utiliser cet outil pour surveiller les actions d'échange de disque.

Goulots d'étranglement d'exécution

Activez les fonctionnalités de performances et réglez les propriétés du service d'intégration de données pour optimiser les performances du mappage. Configurez les paramètres d'optimisation pour le service d'intégration de données et le service de référentiel modèle dans l'outil Administrator.

Allouez de la mémoire pour que les performances du système soient optimales et configurez des niveaux de traçage d'erreur pour réduire le nombre d'événements du journal générés par le service d'intégration de données lorsqu'il exécute le mappage.

Vous pouvez configurer la quantité maximale de mémoire que le service d'intégration de données alloue pour l'exécution de toutes les demandes simultanées. Vous pouvez également limiter la quantité maximale de mémoire que le service d'intégration de données alloue pour toute demande.

Vous pouvez configurer le cache d'ensemble de résultats pour permettre au service d'intégration de données de mettre en cache les résultats du processus DTM associé à chaque requête du service de données SQL et chaque requête du service Web.

Goulots d'étranglement de l'optimisation du service de données SQL

Vous pouvez optimiser les services de données SQL pour améliorer les performances lorsque les utilisateurs finaux exécutent des requêtes SQL sur ces services à l'aide d'outils client tiers. Si un service de données SQL utilise un mappage de table virtuelle, vous pouvez optimiser les transformations et le mappage.

Vous pouvez optimiser le pilote JDBC pour améliorer les performances lors de l'exécution d'une requête sur un service de données SQL. Vous pouvez également configurer le cache d'objet de données pour que le Data Integration Service améliore les performances des mappages et des requêtes SQL.

Goulots d'étranglement de l'optimisation du service Web

Vous pouvez optimiser les services Web pour améliorer les performances lorsque le Data Integration Service exécute des requêtes de service Web. Réglez le Data Integration Service pour qu'il gère la mémoire, traite les requêtes de service Web simultanées et garde un processus DTM actif afin de pouvoir traiter plusieurs requêtes de service Web.

Pour améliorer les performances du service Web, utilisez la compression de message de service Web, optimisez les requêtes HTTP et configurez le cache d'objet de données.

Goulots d'étranglement de la connexion

Vous pouvez optimiser les connexions pour améliorer les performances. Vous pouvez gérer le pool d'instances de connexions inactives pour une connexion de base de données. Vous pouvez augmenter la taille du paquet réseau pour permettre à de plus grands paquets de données de traverser le réseau simultanément.

CHAPITRE 2

Optimisation de cible

Ce chapitre comprend les rubriques suivantes :

- [Présentation de l'optimisation de cible, 16](#)
- [Optimisation de cible de fichier plat, 16](#)
- [Intervalles de point de contrôle de base de données, 17](#)
- [Chargements en bloc, 17](#)
- [Optimisation de cible de base de données, 18](#)

Présentation de l'optimisation de cible

Optimisez les cibles pour permettre au Data Integration Service d'écrire dans les cibles efficacement. Vous pouvez supprimer les index et les contraintes de clé avant d'exécuter un mappage, augmenter le nombre d'intervalles de point de contrôle dans la base de données, configurer le chargement en bloc dans les propriétés d'écriture d'un objet de données et optimiser une base de données cible Oracle.

Utilisez les techniques d'optimisation suivantes pour optimiser la cible :

- Optimisez les cibles de fichiers plats.
- Augmentez les intervalles de point de contrôle de base de données.
- Utilisez les chargements en bloc.
- Optimisez les bases de données cible Oracle.

Optimisation de cible de fichier plat

Vous pouvez améliorer les performances du mappage en optimisant les cibles de fichier plat. Vous pouvez également pousser les tâches de transformation vers une commande pour améliorer les performances.

Envisagez les solutions suivantes pour réduire les goulots d'étranglement de cible de fichier plat :

Poussez les tâches de transformation vers une commande plutôt que vers le Data Integration Service.

Vous pouvez améliorer les performances du mappage en poussant les tâches de transformation vers une commande plutôt que vers le Data Integration Service. Vous pouvez également utiliser une commande pour trier ou compresser les données cible. Dans l'outil Developer, configurez la propriété Commande dans les propriétés d'exécution d'un fichier cible simple.

Sous UNIX, utilisez toute commande ou tout script d'environnement d'exécution UNIX valide. Sous Windows, utilisez toute commande ou tout fichier de commande DOS valide. Le dispositif d'écriture de fichier plat envoie les données à la commande plutôt qu'à une cible de fichier plat.

Par exemple, utilisez la commande suivante pour générer un fichier compressé à partir des données cible :

```
compress -c - > MyTargetFiles/MyCompressedFile.Z
```

Écrivez dans une cible de fichier plat située au niveau du nœud du processus de service.

Si le Data Integration Service fonctionne sur un nœud unique et qu'il écrit dans une cible de fichier plat, vous pouvez optimiser les performances du mappage en écrivant dans une cible de fichier plat située au niveau du nœud du processus de service.

Intervalles de point de contrôle de base de données

Les performances du Data Integration Service diminuent à chaque fois qu'il attend que la base de données effectue un point de contrôle.

Envisagez les solutions suivantes pour réduire les goulots d'étranglement des points de contrôle de base de données :

Augmentez l'intervalle de point de contrôle dans la base de données.

Pour réduire le nombre de points de contrôle et augmenter les performances, augmentez l'intervalle de point de contrôle dans la base de données.

Bien que vous gagniez en performances lorsque vous réduisez le nombre de points de contrôle, vous augmentez aussi le temps de récupération si la base de données s'arrête de façon inattendue.

Chargements en bloc

Lorsque vous utilisez le chargement en bloc, le Data Integration Service contourne le journal de base de données, ce qui augmente les performances.

Envisagez les solutions suivantes pour réduire les goulots d'étranglement du chargement en bloc :

Configurez le chargement en bloc dans les propriétés d'écriture pour un objet de données.

Vous pouvez utiliser le chargement en bloc pour améliorer les performances d'un mappage qui insère une grande quantité de données dans une base de données DB2, Sybase ASE, Oracle ou Microsoft SQL Server.

Sans l'écriture dans le journal de base de données, la base de données cible ne peut pas effectuer de restauration. Par conséquent, vous ne pouvez peut-être pas effectuer de récupération. Lorsque vous utilisez le chargement en bloc, mesurez l'importance de l'amélioration des performances du mappage par rapport à la possibilité de récupérer un mappage incomplet.

Optimisation de cible de base de données

Vous pouvez optimiser la base de données cible en vérifiant la clause de stockage, l'allocation d'espace et les segments de restauration ou d'annulation.

Envisagez les solutions suivantes pour réduire les goulots d'étranglement de cible de base de données :
Vérifiez que la base de données stocke des segments de restauration ou d'annulation dans les espaces de table appropriés, de préférence sur des disques différents.

Lorsque vous écrivez dans la base de données, celle-ci utilise des segments de restauration ou d'annulation pendant les chargements. Demandez à l'administrateur de base de données de veiller à ce que la base de données stocke des segments de restauration ou d'annulation dans les espaces de table appropriés, de préférence sur des disques différents. Les segments de restauration ou d'annulation doivent également comporter les clauses de stockage appropriées.

Réglez le journal de rétablissement de la base de données.

Pour optimiser la base de données, réglez le journal de rétablissement de la base de données. La base de données utilise le journal de rétablissement pour consigner les opérations de chargement. Assurez-vous que la taille du journal de rétablissement et la taille du tampon sont optimales. Pour une base de données Oracle, vous pouvez afficher les propriétés du journal de rétablissement dans le fichier `init.ora`.

Connectez-vous à une base de données Oracle avec le protocole IPC.

Si le Data Integration Service s'exécute sur un nœud unique et que l'instance Oracle est située au niveau du nœud du processus de service, vous pouvez optimiser les performances en utilisant le protocole IPC pour vous connecter à la base de données Oracle. Vous pouvez définir la connexion à la base de données Oracle dans `listener.ora` et `tnsnames.ora`.

CHAPITRE 3

Optimisation de source

Ce chapitre comprend les rubriques suivantes :

- [Présentation de l'optimisation de source, 19](#)
- [Optimisation de source de fichier plat, 20](#)
- [Optimisation de requête, 20](#)
- [Filtres conditionnels, 21](#)
- [Sélectionner distinct, 21](#)
- [Indicateurs, 22](#)
- [Contraintes, 23](#)
- [Optimisation d'objet de données personnalisé, 24](#)
- [Optimisation de source de base de données, 25](#)

Présentation de l'optimisation de source

Optimisez les sources de fichier plat, les sources relationnelles et de données personnalisées pour permettre au service d'intégration de données de lire les données source de façon efficace.

Utilisez les techniques d'optimisation suivantes pour optimiser les sources :

- Lisez les données source de façon efficace.
- Utilisez les techniques d'optimisation de requête.
- Utilisez des filtres conditionnels avec la requête SQL.
- Sélectionnez des valeurs uniques à partir de la source.
- Appliquez des indicateurs à la requête SQL.
- Configurez les contraintes sur les objets de données logiques, les objets de données physiques, et les tables virtuelles.
- Configurez l'optimisation des objets de données personnalisés.
- Configurez l'optimisation des bases de données Oracle, Sybase et Microsoft SQL Server.

Optimisation de source de fichier plat

Configurez les propriétés de format des sources de fichier plat pour permettre au Data Integration Service de lire les données source de façon efficace.

Envisagez les solutions suivantes pour les goulots d'étranglement de source de fichier plat :

N'utilisez pas de guillemets ou de caractères d'échappement dans les propriétés de format d'un fichier plat délimité.

Si vous spécifiez un caractère d'échappement, le Data Integration Service lit le caractère délimiteur comme un caractère normal intégré dans la chaîne. Vous pouvez améliorer légèrement les performances du mappage si le fichier source ne contient pas de guillemets ou de caractères d'échappement.

Définissez le nombre d'octets que le Data Integration Service lit par ligne.

Si le mappage lit à partir d'une source de fichier plat, vous pouvez améliorer les performances du mappage en définissant le nombre d'octets que le Data Integration Service lit par ligne. Configurez la propriété Longueur de tampon de la ligne séquentielle dans les propriétés d'exécution des sources de fichier plat.

Par défaut, le Data Integration Service lit 1024 octets par ligne. Si chaque ligne dans le fichier source comporte un nombre d'octets inférieur au paramétrage par défaut, vous pouvez réduire la longueur de tampon de la ligne séquentielle dans les propriétés de mappage.

Optimisation de requête

Si un mappage joint plusieurs tables source dans un objet de données personnalisé, vous pourrez peut-être améliorer les performances en optimisant la requête avec des indicateurs d'optimisation. De même, les instructions Select de table unique avec une clause ORDER BY ou GROUP BY peuvent bénéficier de l'optimisation, par exemple en ajoutant des index.

Envisagez les solutions suivantes pour les goulots d'étranglement de requête :

Créez des indicateurs d'optimisation pour indiquer à la base de données comment exécuter la requête pour un ensemble de tables source particulier.

Généralement, l'optimiseur de base de données détermine la manière la plus efficace de traiter les données source. Cependant, vous connaissez peut-être des propriétés des tables source que l'optimiseur de base de données ne connaît pas. L'administrateur de base de données peut créer des indicateurs d'optimisation pour indiquer à la base de données comment exécuter la requête pour un ensemble de tables source particulier.

Configurez des indicateurs d'optimisation pour commencer à renvoyer des lignes le plus rapidement possible, plutôt que de renvoyer toutes les lignes à la fois.

Utilisez des indicateurs d'optimisation s'il y a un long délai entre le moment où la requête commence à s'exécuter et le moment où le Data Integration Service reçoit la première ligne de données. Configurez des indicateurs d'optimisation pour commencer à renvoyer des lignes le plus rapidement possible, plutôt que de renvoyer toutes les lignes à la fois. Cela permet au Data Integration Service de traiter des lignes parallèlement à l'exécution de la requête.

Créez un index sur les colonnes ORDER BY ou GROUP BY.

Les requêtes contenant des clauses ORDER BY ou GROUP BY peuvent bénéficier de la création d'un index sur les colonnes ORDER BY ou GROUP BY. Une fois que vous avez optimisé la requête, utilisez l'option de remplacement SQL pour profiter pleinement de ces modifications.

Configurez la base de données pour exécuter les requêtes parallèles.

Vous pouvez également configurer la base de données source pour exécuter des requêtes parallèles afin d'améliorer les performances. Pour plus d'informations sur la configuration de requêtes parallèles, consultez la documentation de la base de données.

La requête que le Data Integration Service utilise pour lire les données apparaît dans la base de données virtuelle dans un service de données SQL. Vous pouvez également trouver la requête dans l'objet de données personnalisé. Demandez à l'administrateur de base de données d'analyser la requête, puis créez des indicateurs d'optimisation et des index pour les tables sources.

La requête que le service d'intégration de données utilise pour lire les données s'affiche dans l'objet de données personnalisé. Demandez à l'administrateur de base de données d'analyser la requête, puis créez des indicateurs d'optimisation et des index pour les tables sources.

Filtres conditionnels

Un filtre source simple sur la base de données source peut parfois avoir un impact négatif sur les performances en raison du manque d'index. Vous pouvez utiliser le filtre conditionnel dans l'objet de données personnalisé pour augmenter les performances.

Envisagez la solution suivante pour les goulots d'étranglement de filtre conditionnel :

Utilisez le filtre conditionnel pour plusieurs mappages lisant simultanément dans la même source.

Si plusieurs mappages lisent dans la même source simultanément, le filtre conditionnel peut améliorer les performances.

Pendant, certains mappages peuvent s'effectuer plus rapidement si vous filtrez les données source sur la base de données source. Vous pouvez tester le mappage avec le filtre de base de données et le filtre conditionnel en même temps pour déterminer quelle méthode améliore les performances.

Sélectionner distinct

Vous pouvez sélectionner des valeurs uniques à partir de sources dans un objet de données personnalisé avec l'option Sélectionner distinct. Lorsque vous utilisez Sélectionner distinct, le Data Integration Service ajoute une instruction SELECT DISTINCT à la requête SQL par défaut.

Envisagez la solution suivante pour les goulots d'étranglement de Sélectionner distinct :

Utilisez l'option Sélectionner distinct pour filtrer des données inutiles plus tôt dans le flux de données.

Utilisez l'option Sélectionner distinct pour l'objet de données personnalisé si vous souhaitez que le Data Integration Service sélectionne des valeurs uniques à partir d'une source. Utilisez l'option Sélectionner distinct pour filtrer des données inutiles plus tôt dans le flux de données. Cela peut améliorer les performances.

Par exemple, vous pouvez utiliser l'option Sélectionner distinct pour extraire des ID de client uniques à partir d'une table qui répertorie des ventes totales. Lorsque vous utilisez l'objet de données personnalisé dans un mappage, le Data Integration Service filtre les données inutiles plus tôt dans le flux de données, ce qui peut augmenter les performances.

Indicateurs

Vous pouvez ajouter des indicateurs à la requête SQL source pour transmettre des instructions à un optimiseur de base de données. L'optimiseur utilise ces indicateurs pour choisir un plan d'exécution de requête pour accéder à la source.

Le champ Indicateurs apparaît dans la vue **Requête** d'une instance d'objet de données relationnel ou d'un objet de données personnalisé. La base de données source doit être une base de données Oracle, Sybase, IBM DB2 ou Microsoft SQL Server. Le champ Indicateurs n'apparaît pas pour les autres types de base de données.

Lorsque le Data Integration Service génère la requête source, il ajoute les indicateurs SQL à la requête exactement comme vous les saisissez dans l'outil Developer. Le Data Integration Service n'analyse pas les indicateurs. Lorsque vous exécutez le mappage qui contient la source, le journal de mappage montre la requête avec les indicateurs dans la requête.

Le Data Integration Service insère les indicateurs SQL dans un emplacement dans la requête selon le type de base de données. Reportez-vous à la documentation de votre base de données pour connaître la syntaxe pour les indicateurs.

Oracle

Le Data Integration Service ajoute des indicateurs directement après le mot-clé SELECT/UPDATE/INSERT/DELETE.

```
SELECT /*+ <indicateurs> */ FROM ...
```

Le « + » indique le début des indicateurs.

Les indicateurs sont contenus dans un commentaire (/ * ... */ ou --... jusqu'à la fin de la ligne)

Sybase

Le Data Integration Service ajoute des indicateurs après la requête. Configurez un nom de plan dans l'indicateur.

```
SELECT ... PLAN <plan>
```

```
select avg(price) from titles plan « (scalar_agg (i_scan type_price_ix titles ) »
```

IBM DB2

Vous pouvez saisir la clause optimize-for comme indicateur. Le Data Integration Service ajoute la clause à la fin de la requête.

```
SELECT ... OPTIMIZE FOR <n> ROWS
```

La clause optimize-for indique à l'optimiseur de base de données combien de lignes la requête pourra traiter. La clause ne limite pas le nombre de lignes. Si la base de données traite plus de <n> lignes, les performances risquent de diminuer.

Microsoft SQL Server

Le Data Integration Service ajoute des indicateurs à la fin de la requête dans le cadre d'une clause OPTION.

```
SELECT ... OPTION ( <query_hints> )
```

Règles et instructions concernant les indicateurs

Utilisez les règles et instructions suivantes lorsque vous configurez des indicateurs pour les requêtes SQL :

- Si vous activez l'optimisation de refoulement ou si vous utilisez une semi-jointure dans un objet de données relationnel, la requête source d'origine change. Le Data Integration Service n'applique pas d'indicateurs à la requête modifiée.
- Vous pouvez combiner des indicateurs avec des remplacements de jointure et de filtre, mais si vous configurez un remplacement SQL, celui-ci est prioritaire et le Data Integration Service n'applique pas les autres remplacements.
- La vue **Requête** affiche une vue simple ou avancée. Si vous saisissez un indicateur avec un remplacement de filtre, de tri ou de jointure sur la vue simple, l'outil Developer affiche la totalité du remplacement de requête dans la vue avancée.

Création d'indicateurs

Créez des indicateurs pour envoyer des instructions à l'optimiseur de base de données pour déterminer un plan de requête.

1. Ouvrez l'instance d'objet de données personnalisé ou d'objet de données relationnel.
2. Sélectionnez la vue **Lecture**.
3. Sélectionnez la transformation Sortie.
4. Sélectionnez les propriétés de **Requête**.
5. Sélectionnez la requête simple.
6. Cliquez sur **Éditer** en regard du champ **Indicateurs**.
La boîte de dialogue **Indicateurs** s'affiche.
7. Saisissez l'indicateur dans le champ **Requête SQL**.
L'outil Developer ne valide pas l'indicateur.
8. Cliquez sur **OK**.
9. Enregistrez l'objet de données.

Contraintes

Le service d'intégration de données peut lire les contraintes à partir des sources relationnelles, des sources de fichiers plats, des objets de données logiques ou des tables virtuelles. Une contrainte est une expression conditionnelle à laquelle doivent satisfaire les valeurs d'une ligne de données.

Lorsque le service d'intégration de données lit les contraintes, il est possible qu'il annule les lignes de données n'ayant pas la valeur True en fonction de la méthode d'optimisation appliquée.

Avant de définir une contrainte, vous devez vérifier que les données source sont conformes à la condition définie par la contrainte.

Par exemple, une base de données source a une colonne AGE dont certaines lignes contiennent AGE < 70. Vous pouvez définir une contrainte AGE < 70 sur la base de données source. L'intégration de données lit les enregistrements de la base de données source avec la contrainte âge < 70. Si le service d'intégration de données lit les enregistrements contenant AGE > = 70, il est possible qu'il annule les lignes affichant AGE > = 70.

Dans la base de données, vous pouvez utiliser des commandes SQL pour définir des contraintes sur l'environnement de la base de données lorsque vous vous connectez à la base de données. Le service

d'intégration de données exécute l'environnement SQL de connexion à chaque connexion à la base de données.

Utilisez l'outil Developer pour définir les contraintes sur les objets de données logiques, les objets de données physiques, et les tables virtuelles. Lorsque vous définissez une contrainte, vous devez entrer une expression qui renvoie la valeur TRUE pour chaque ligne de données.

Configuration de contraintes

Vous pouvez ajouter des contraintes à des objets de données relationnels, des objets de données de fichier plat, des objets de données personnalisés, des objets de données logiques et des tables virtuelles. Après avoir ajouté une contrainte, vous pouvez la modifier ou la supprimer.

1. Dans la vue **Explorateur d'objets**, ouvrez le mappage qui contient l'objet de données relationnel ajouté en tant que transformation Lecture. Vous pouvez également ouvrir l'objet de données de fichier plat, un objet de données personnalisé, un objet de données logique ou une table virtuelle.
 - Pour définir les contraintes sur un objet de données relationnel ajouté à un mappage en tant que transformation Lecture, sélectionnez la transformation Lecture dans le mappage. Dans la vue **Propriétés**, sélectionnez l'onglet **Avancé**.
 - Pour définir les contraintes sur un objet de données de fichier plat, sélectionnez la vue **Avancé** et développez la section **Exécution : lecture** section.
 - Pour définir des contraintes sur un objet de données personnalisé, sélectionnez la vue **Lire**, puis le port **Sortie** de la transformation source. Dans la vue **Propriétés**, sélectionnez l'onglet **Avancé**.
 - Pour définir des contraintes sur un objet de données logique, sélectionnez un modèle de données logique et l'objet de données logique. Dans la vue **Propriétés**, sélectionnez l'onglet **Avancé**.
 - Pour définir des contraintes sur une table virtuelle, ouvrez la table virtuelle depuis le point de terminaison SQL. Dans la vue **Propriétés**, sélectionnez l'onglet **Avancé**.
2. Cliquez sur le champ de valeur des contraintes.
La boîte de dialogue **Contraintes** s'ouvre.
3. Cliquez sur **Nouveau** pour ouvrir l'éditeur d'expression.
4. Configurez la logique de contrainte et utilisez les fonctions d'expression et les colonnes comme paramètres.
5. Cliquez sur **Valider**.
6. Cliquez sur **OK**.

Optimisation d'objet de données personnalisé

Vous pouvez configurer des objets de données personnalisés pour améliorer les performances. Vous pouvez optimiser la requête SQL, utiliser des filtres conditionnels et sélectionner des valeurs distinctes à partir de la source dans un objet de données personnalisé.

Envisagez les solutions suivantes pour les goulots d'étranglement d'objets de données personnalisés :

Créez une requête personnalisée pour émettre une instruction SELECT spéciale pour que le service d'intégration de données lise les données source.

La requête personnalisée remplace la requête par défaut que le service d'intégration de données utilise pour lire les données à partir des sources.

Filtrez les lignes lorsque le service d'intégration de données lit les données source.

Si vous incluez une condition de filtre, le service d'intégration de données ajoute une clause WHERE à la requête par défaut.

Sélectionnez des valeurs distinctes à partir de la source.

Si vous avez choisi Sélectionner distinct, le service d'intégration de données ajoute une instruction SELECT DISTINCT à la requête SQL par défaut.

Appliquez des indicateurs de base de données.

Vous pouvez ajouter des indicateurs à la requête SQL source pour transmettre des instructions à un optimiseur de base de données.

Configurez les contraintes sur les données source.

Si vous configurez les contraintes sur les fichiers plats et les tables relationnelles dans un objet de données personnalisé, le service d'intégration de données annule les lignes qui ne renvoient pas la valeur TRUE pour les lignes de données.

Optimisation de source de base de données

Si la base de données source est Oracle, vous pouvez optimiser les performances du Data Integration Service en utilisant le protocole IPC pour vous connecter à la base de données Oracle. Vous pouvez également déplacer la base de données temporaire vers une baie de disques pour améliorer les performances.

Envisagez les solutions suivantes pour les goulots d'étranglement de source de base de données :

Utilisez le protocole IPC pour vous connecter à la base de données Oracle.

Si le Data Integration Service s'exécute sur un nœud unique et que l'instance Oracle est située au niveau du nœud du processus de service, vous pouvez optimiser les performances en utilisant le protocole IPC pour vous connecter à la base de données Oracle. Vous pouvez définir la connexion à la base de données Oracle dans listener.ora et tnsnames.ora.

Déplacez la base de données temporaire et recréez les journaux sur une baie de disques ou sur des disques plus rapides.

Lorsque vous joignez de grandes tables dans une base de données, vous pouvez utiliser une baie redondante de disques indépendants (RAID) pour l'emplacement du cache. Vous pouvez également ajouter des fichiers vers le groupe de fichiers principal sur d'autres disques pour répartir la charge entre les disques.

CHAPITRE 4

Optimisation de transformation

Ce chapitre comprend les rubriques suivantes :

- [Optimisation de transformation, 26](#)
- [Optimisation de la transformation Agrégation, 27](#)
- [Transformation Processeur de données, 27](#)
- [Optimisation d'expression, 28](#)
- [Optimisation de la transformation Java, 30](#)
- [Optimisation de la transformation Jointure, 32](#)
- [Optimisation de la transformation Recherche, 33](#)
- [Optimisation de la transformation Trieur, 36](#)
- [Optimisation de la transformation SQL, 36](#)
- [Cache de transformation, 38](#)
- [Élimination d'erreurs de transformation, 39](#)
- [Effets secondaires de transformation, 39](#)
- [Optimisation de la transformation Consommateur de service Web, 40](#)

Optimisation de transformation

Optimisez les transformations pour permettre au Data Integration Service de traiter des transformations dans un mappage de façon efficace.

Utilisez les techniques d'optimisation suivantes pour optimiser la transformation :

- Configurez l'optimisation pour les transformations.
- Éliminez les erreurs de transformation.
- Configurez le cache de transformation.

Optimisation de la transformation Agrégation

Les transformations Agrégation ralentissent souvent les performances car elles doivent regrouper des données avant de les traiter. Les transformations Agrégation ont besoin de mémoire supplémentaire pour conserver les résultats de groupe intermédiaire.

Envisagez les solutions suivantes pour les goulots d'étranglement de la transformation Agrégation :

Regroupez par colonnes simples.

Vous pouvez optimiser les transformations Agrégation lorsque vous regroupez par colonnes simples. Lorsque cela est possible, utilisez des nombres plutôt que des chaînes et des dates dans les colonnes utilisées pour GROUP BY. Évitez les expressions complexes dans les expressions Agrégation.

Utilisez une entrée triée.

Pour augmenter les performances de mappage, trie les données pour la transformation Agrégation. Utilisez l'option Entrée triée pour trier les données.

L'option Entrée triée réduit l'utilisation de caches d'agrégation. Lorsque vous utilisez l'option Entrée triée, le Data Integration Service considère que toutes les données sont triées par groupe. Lorsque le Data Integration Service lit les lignes pour un groupe, il effectue des calculs d'agrégation. Au besoin, il stocke des informations de groupe en mémoire.

L'option Entrée triée réduit la quantité de données mises en cache pendant le mappage et améliore les performances. Utilisez l'option Entrée triée ou une transformation Trieur pour transmettre les données triées à la transformation Agrégation.

Vous pouvez augmenter les performances lorsque vous utilisez l'option Entrée triée dans les mappages avec plusieurs partitions.

Filtrez les données avant de les regrouper.

Si vous utilisez une transformation Filtre dans le mappage, placez la transformation avant la transformation Agrégation pour réduire les agrégations inutiles.

Limitez les connexions de ports.

Limitez le nombre de ports d'entrées/sorties ou de sortie connectés pour réduire la quantité de données stockée par la transformation Agrégation dans le cache de données.

Transformation Processeur de données

Les transformations Processeur de données peuvent réduire les performances lorsque le service d'intégration de données passe une seule ligne à la procédure de transformation au lieu d'un bloc de lignes dans un tableau.

Envisagez les solutions suivantes pour les goulots d'étranglement liés aux transformations Processeur de données :

Écrivez le code de procédure pour spécifier que la procédure reçoit un bloc de lignes.

Vous pouvez augmenter les performances lorsque la procédure reçoit un bloc de lignes.

Réduisez le nombre d'appels de fonction effectués par le service d'intégration de données et la procédure.

Le service d'intégration de données appelle moins de fois la fonction de notification de la ligne d'entrée et la procédure appelle moins de fois la fonction de notification de sortie.

Augmentez la localité des accès à la mémoire.

Les performances du service d'intégration de données peuvent être optimisées si vous augmentez la localité de l'accès à la mémoire pour les données.

Écrivez le code de procédure pour exécuter un algorithme sur un bloc de données plutôt que sur chaque ligne de données.

Vous pouvez augmenter les performances lorsque la procédure exécute un algorithme sur un bloc de données.

Optimisation d'expression

Certaines expressions utilisées dans une transformation peuvent réduire les performances.

Envisagez les solutions suivantes pour les goulots d'étranglement d'expression :

Isolez les expressions lentes.

Les expressions lentes ralentissent les performances du mappage. Pour isoler les expressions lentes, retirez-les du mappage une par une, puis exécutez le mappage pour déterminer la durée d'exécution du mappage sans l'expression. Si la différence d'exécution du mappage est importante, cherchez des moyens d'optimiser l'expression lente.

Procédez comme suit pour évaluer les performances de l'expression :

1. Chronométrez le mappage avec les expressions d'origine.
2. Copiez le mappage et remplacez la moitié des expressions complexes par une constante.
3. Exécutez et chronométrez le mappage modifié.
4. Faites une autre copie du mappage et remplacez l'autre moitié des expressions complexes par une constante.
5. Exécutez et chronométrez le mappage modifié.

Excluez la logique commune.

Si le mappage effectue la même tâche à plusieurs endroits, réduisez le nombre de fois que le mappage effectue la tâche en déplaçant la tâche plus tôt dans le mappage. Par exemple, vous avez un mappage avec cinq tables cible. Chaque cible requiert une recherche de numéro de sécurité sociale. Au lieu d'effectuer la recherche cinq fois, placez la transformation Recherche dans le mappage avant que le flux de données ne se sépare. Ensuite, transmettez les résultats de la recherche aux cinq cibles.

Limitez les appels de la fonction d'agrégation.

Lors de l'écriture d'expressions, excluez autant d'appels de la fonction d'agrégation que possible. À chaque fois que vous utilisez un appel de la fonction d'agrégation, le service d'intégration de données doit rechercher et regrouper les données. Par exemple, dans l'expression suivante, le service d'intégration de données lit COLUMN_A, trouve la somme, puis lit COLUMN_B, trouve la somme, puis trouve la somme des deux sommes :

```
SUM(COLUMN_A) + SUM(COLUMN_B)
```

Si vous excluez l'appel de la fonction d'agrégation, comme ci-dessous, le service d'intégration de données ajoute COLUMN_A à COLUMN_B, puis trouve la somme des deux.

```
SUM(COLUMN_A + COLUMN_B)
```

Remplacez les expressions communes par des variables locales.

Si vous utilisez la même expression plusieurs fois dans une transformation, vous pouvez faire de cette expression une variable locale. Vous ne pouvez utiliser une variable locale qu'à l'intérieur de la transformation. Cependant, en ne calculant la variable qu'une seule fois, vous augmentez les performances.

Choisissez des opérateurs numériques plutôt que des opérateurs de chaînes.

Le service d'intégration de données traite les opérations numériques plus vite que les opérations de chaînes. Par exemple, si vous recherchez de grandes quantités de données sur deux colonnes, EMPLOYEE_NAME et EMPLOYEE_ID, en configurant la recherche autour de EMPLOYEE_ID, vous améliorez les performances.

Optimisez les comparaisons CHAR-CHAR et CHAR-VARCHAR.

Lorsque le service d'intégration de données effectue des comparaisons entre les colonnes CHAR et VARCHAR, il ralentit dès qu'il trouve des espaces de fin dans la ligne. Vous pouvez utiliser l'option TreatCHARasCHARonRead quand vous configurez le service d'intégration de données dans Informatica Administrator afin que celui-ci ne supprime pas les espaces à la fin des champs source Char.

Choisissez DECODE plutôt que LOOKUP.

Lorsque vous utilisez une fonction LOOKUP, le service d'intégration de données doit rechercher une table dans une base de données. Lorsque vous utilisez une fonction DECODE, vous intégrez les valeurs de recherche dans l'expression pour que le service d'intégration de données n'ait pas à rechercher une table distincte. Par conséquent, lorsque vous voulez rechercher un petit ensemble de valeurs qui ne changent pas, utilisez DECODE pour améliorer les performances.

Utilisez des opérateurs plutôt que des fonctions.

Le service d'intégration de données lit les expressions écrites avec des opérateurs plus rapidement que les expressions avec des fonctions. Lorsque vous le pouvez, utilisez des opérateurs pour écrire des expressions. Par exemple, l'expression suivante contient des fonctions CONCAT imbriquées :

```
CONCAT( CONCAT( CUSTOMERS.FIRST_NAME, ' ') CUSTOMERS.LAST_NAME)
```

Vous pouvez réécrire cette expression avec l'opérateur || comme suit :

```
CUSTOMERS.FIRST_NAME || ' ' || CUSTOMERS.LAST_NAME
```

Optimisez les fonctions IIF.

Les fonctions IIF peuvent renvoyer une valeur et une action, ce qui permet des expressions plus compactes. Par exemple, une source comporte trois indicateurs Y/N : FLG_A, FLG_B, FLG_C. Vous souhaitez renvoyer des valeurs basées sur celles de chaque indicateur.

Vous utilisez l'expression suivante :

```
IIF( FLG_A = 'Y' and FLG_B = 'Y' AND FLG_C = 'Y',  
VAL_A + VAL_B + VAL_C,  
IIF( FLG_A = 'Y' and FLG_B = 'Y' AND FLG_C = 'N',  
VAL_A + VAL_B ,  
IIF( FLG_A = 'Y' and FLG_B = 'N' AND FLG_C = 'Y',  
VAL_A + VAL_C,  
IIF( FLG_A = 'Y' and FLG_B = 'N' AND FLG_C = 'N',  
VAL_A ,  
IIF( FLG_A = 'N' and FLG_B = 'Y' AND FLG_C = 'Y',  
VAL_B + VAL_C,  
IIF( FLG_A = 'N' and FLG_B = 'Y' AND FLG_C = 'N',  
VAL_B ,  
IIF( FLG_A = 'N' and FLG_B = 'N' AND FLG_C = 'Y',  
VAL_C,  
IIF( FLG_A = 'N' and FLG_B = 'N' AND FLG_C = 'N',  
0.0,  
))))))
```

Optimisation de la transformation Java

Certaines transformations Java dans un mappage peuvent réduire les performances.

Envisagez les solutions suivantes pour augmenter les performances de la transformation Java :

Activez la sélection précoce ou les méthodes d'optimisation de filtre push-into avec la transformation Java.

Vous pouvez activer la sélection précoce ou l'optimisation push-into dans les transformations Java. Mettez à jour les snippets de code dans l'onglet **Interfaces optimiseur** de la transformation Java.

Optimisation de sélection précoce avec la transformation Java

Vous pouvez activer une transformation Java active ou passive pour l'optimisation de sélection précoce si la transformation Java n'a pas d'effets secondaires. L'optimiseur transmet la logique de filtre via la transformation Java et modifie la condition de filtre comme requis.

Pour afficher les snippets de code pour l'optimisation de sélection précoce, choisissez PredicatePushOptimization dans le navigateur de l'onglet **Interfaces optimiseur**.

allowPredicatePush

Booléen. Active la sélection précoce. Modifiez la fonction pour qu'elle renvoie un résultat True et un message pour activer la sélection précoce. La valeur par défaut est False et la fonction renvoie un message comme quoi l'optimisation n'est pas prise en charge.

```
public ResultAndMessage allowPredicatePush(boolean ignoreOrderOfOp) {
    // To Enable PredicatePushOptimization, this function should return true
    //return new ResultAndMessage(true, "");
    return new ResultAndMessage(false, "Predicate Push Optimization Is Not
Supported");
}
```

canGenerateOutputFieldEvalError

Booléen. Indique si la transformation Java peut renvoyer ou non une erreur de champ de sortie, telle qu'une erreur de division par zéro. Modifiez la fonction pour qu'elle renvoie False si la transformation Java ne génère pas d'erreurs de champ de sortie. Quand la transformation Java peut générer des erreurs de champ, le service d'intégration de données ne peut utiliser l'optimisation de sélection précoce.

```
public boolean canGenerateOutputFieldEvalError() {
    // If this Java transformation can never generate an output field evaluation error,
    // return false.
    return true;
}
```

getInputExpr

Renvoie une expression Informatica qui décrit quelles sont les valeurs d'entrée des champs d'entrée qui comprennent un champ de sortie. L'optimiseur doit savoir quels sont les champs d'entrée qui comprennent un champ de sortie pour pousser la logique de filtre via la transformation.

```
public InfaExpression getInputExpr(TransformationField field,
TransformationDataInterface group) {
    // This should return an Informatica expression for output fields in terms of input
fields
    // We will only push predicate that use fields for which input expressions are
defined.
    // For example, if you have two input fields in0 and in1 and three output fields
out0, out1, out2
    // out0 is the pass-through of in1, out2 is sum of in1 and in2, and out3 is unknown,
the code should be:
    //if (field.getName().equals("out0"))
    // return new InfaExpression("in0", instance);
}
```

```

        //else if (field.getName().equals("out1"))
        //    return new InfaExpression("in0 + in1", instance);
        //else if (field.getName().equals("out2"))
        //    return null;
        return null;
    }
}

```

Par exemple : un mappage contient l'expression de filtre « out0 > 8 ». Out0 est la valeur du port de sortie out0 dans la transformation Java. Vous pouvez définir la valeur de out0 en tant que valeur du port d'entrée in0 + 5. L'optimiseur peut pousser l'expression suivante « (in0 + 5) > 8 » après la transformation Java avec l'optimisation de sélection précoce. Vous pouvez renvoyer NULL si un champ de sortie n'a pas l'expression du champ d'entrée. L'optimiseur ne pousse pas les expressions de filtre après les champs de sortie sans expression d'entrée.

Vous pouvez inclure le code suivant :

```

    if (field.getName().equals("out0"))
        return new InfaExpression("in0 + 5", instance);
    else if (field.getName().equals("out2"))
        return null;
}

```

inputGroupsPushPredicateTo

Revoie une liste de groupes pouvant recevoir la logique de filtre. La transformation Java a un groupe d'entrée. Ne modifiez cette fonction de la transformation Java.

```

public List<TransformationDataInterface> inputGroupsPushPredicateTo(
    List<TransformationField> fields) {
    // This functions returns a list of input data interfaces to push predicates to.
    // Since JavaTx only has one input data interface, you should not have to modify
    this function
    AbstractTransformation tx = instance.getTransformation();
    List<DataInterface> dis = tx.getDataInterfaces();
    List<TransformationDataInterface> inputDIs = new
    ArrayList<TransformationDataInterface>();
    for (DataInterface di : dis){
        TransformationDataInterface tdi = (TransformationDataInterface) di;
        if (tdi.isInput())
            inputDIs.add(tdi);
    }
    if(inputDIs.size() == 1)
        return inputDIs;
    else
        return null;
}

```

Optimisation push-into avec la transformation Java

Vous pouvez activer une transformation Java active pour l'optimisation push-into si elle n'a pas d'effets secondaires et si l'optimisation n'affecte pas les résultats de mappage.

Lorsque vous configurez l'optimisation push-into pour la transformation Java, vous définissez un moyen pour la transformation Java de stocker la condition de filtre qu'elle reçoit de l'optimiseur. Ajoutez le code qui examine la condition de filtre. Si la transformation Java peut absorber la logique de filtre, la transformation Java transmet alors une condition True à l'optimiseur. L'optimiseur supprime la transformation filtre du mappage optimisé.

Lorsque vous configurez la transformation Java, vous écrivez le code qui stocke la condition de filtre en tant que métadonnées de la transformation lors de l'optimisation. Vous écrivez également le code pour récupérer la condition de filtre lors de l'exécution et pour supprimer les lignes selon la logique de filtre.

Lorsque vous définissez la transformation Java, vous ajoutez du code pour l'optimisation push-into dans l'onglet de la transformation Java **Interfaces optimiseur**. Pour accéder aux snippets de code pour l'optimisation push-into, choisissez FilterPushdownOptimization dans le navigateur de l'onglet de la transformation **Interfaces optimiseur**.

L'outil Developer affiche les snippets de code pour activer l'optimisation push-into et pour recevoir la condition de filtre depuis l'optimiseur. Mettez à jour les snippets de code pour activer l'optimisation et pour enregistrer la logique de filtre en tant que métadonnées de la transformation.

isFilterSupported

Renvoie True pour activer l'optimisation push-into. Renvoie False pour désactiver l'optimisation push-into. Modifiez la fonction de façon à ce qu'elle renvoie True pour activer l'optimisation push-into.

```
public ResultAndMessage isFilterSupported() {
    // To enable filter push-into optimization this function should return true
    // return new ResultAndMessage(true, "");
    return new ResultAndMessage(false, "Filter push-into optimization is not supported");
}
```

pushFilter

Reçoit la condition de filtre de l'optimiseur.

Ajoutez le code pour examiner le filtre et déterminer si la logique de filtre peut être utilisée dans la transformation. Si la transformation peut absorber le filtre, utilisez alors la méthode suivante pour stocker la condition de filtre en tant que métadonnées de transformation :

```
storeMetadata(String key, String data)
```

La clé est un identifiant pour les métadonnées. Vous pouvez définir n'importe quelle chaîne en tant que clé. Les données sont celles que vous voulez stocker afin de déterminer quelles sont les lignes à retirer lors de l'exécution. Par exemple : les données peuvent être la condition de filtre que la transformation Java reçoit de l'optimiseur.

```
public ResultAndMessage pushFilter(InfaExpression condition) {
    // Add code to absorb the filter
    // If filter is successfully absorbed return new ResultAndMessage(true, ""); and the
    optimizer
    // will remove the filter from the mapping
    // If the filter is not absorbed, return new ResultAndMessage(false, msg);
    return new ResultAndMessage(false, "Filter push-into optimization is not supported");
}
```

Optimisation de la transformation Jointure

Les transformations Jointure peuvent ralentir les performances, car elles requièrent un espace supplémentaire lors de l'exécution pour contenir les résultats intermédiaires.

Envisagez les solutions suivantes pour les goulots d'étranglement de la transformation Jointure :

Indiquez la source principale comme source contenant le moins de valeurs de clés dupliquées.

Lorsque le service d'intégration de données traite une transformation Jointure triée, il met en cache les lignes de cent clés uniques à la fois. Si la source principale contient plusieurs lignes avec la même valeur de clé, le service d'intégration de données doit mettre en cache plusieurs lignes, ce qui peut diminuer les performances.

Indiquez la source principale comme source contenant le moins de lignes.

La transformation Jointure compare chaque ligne de la source secondaire avec la source principale. Moins la source principale contient de lignes, moins d'itérations de la comparaison de jointure devront avoir lieu, ce qui accélère la jointure.

Effectuez les jointures dans une base de données lorsque c'est possible.

L'exécution d'une jointure dans une base de données est plus rapide que l'exécution d'une jointure lors de l'exécution du mappage. Le type de jointure de base de données que vous utilisez peut affecter les performances. Les jointures normales sont plus rapides que les jointures extérieures et produisent moins de lignes. Parfois, vous ne pouvez pas effectuer la jointure dans la base de données, par exemple des jointures de tables à partir de deux bases de données ou systèmes de fichiers simples différents.

Joignez des données triées lorsque c'est possible.

Configurez la transformation Jointure pour utiliser l'entrée triée. Le service d'intégration de données améliore les performances en réduisant l'entrée et la sortie de disque. La plus forte augmentation des performances se produit lorsque vous travaillez avec des ensembles de données volumineux. Pour une transformation Jointure non triée, indiquez la source contenant le moins de lignes comme source principale.

Optimisez la condition de jointure.

Le Data Integration Service tente de réduire la taille de l'ensemble de données d'un opérande de jointure en lisant les lignes du plus petit groupe, en trouvant les lignes correspondantes dans le plus grand groupe, et en effectuant ensuite l'opération de jointure. Le fait de réduire la taille de l'ensemble de données améliore les performances du mappage car le Data Integration Service ne lit plus de lignes inutiles dans la source du plus grand groupe. Le Data Integration Service déplace la condition de jointure vers la source du plus grand groupe et ne lit que les lignes qui correspondent au plus petit groupe.

Utilisez la méthode d'optimisation de semi-jointure.

Utilisez la méthode d'optimisation de semi-jointure pour améliorer les performances du mappage lorsqu'un groupe d'entrée a beaucoup plus de lignes que l'autre et lorsque le plus grand groupe a beaucoup de lignes sans correspondance dans le plus petit groupe en fonction de la condition de jointure.

Optimisation de la transformation Recherche

Les transformations Recherche peuvent ralentir les performances en fonction du type de cache de recherche et des conditions de recherche.

Envisagez les solutions suivantes pour les goulots d'étranglement de transformation Recherche :

Utilisez le pilote de base de données optimal.

Le Data Integration Service peut se connecter à une table de recherche à l'aide d'un pilote de base de données natif ou d'un pilote ODBC. Les pilotes de base de données natifs fournissent de meilleures performances de mappage que les pilotes ODBC.

Mettez les tables de recherche en cache pour les recherches de fichiers plats ou relationnelles.

Pour améliorer les performances des recherches dans les fichiers plats et les sources relationnelles, activez la mise en cache des recherches dans la transformation. Lorsque vous activez la mise en cache, le service d'intégration de données met la table de recherche en cache. Lorsque vous exécutez le mappage, le service d'intégration de données interroge le cache de recherche plutôt que la table de recherche. Lorsque cette option n'est pas activée, le Data Integration Service interroge la table de recherche ligne par ligne.

Le résultat de la requête de recherche et le traitement sont les mêmes, que vous mettiez la table de recherche en cache ou non. Cependant, l'utilisation d'un cache de recherche peut augmenter les performances de mappage pour des tables de recherche plus petites. Généralement, vous souhaitez mettre en cache des tables de recherche qui nécessitent moins de 300 Mo.

Mettez les tables de recherche en cache pour les recherches d'objets de données logiques.

Pour améliorer les performances de recherche dans un objet de données logique, vous pouvez activer la mise en cache de l'objet de données sur le service d'intégration de données. Lorsque vous activez la mise en cache de l'objet de données, le service d'intégration de données met l'objet de données logique en cache. Pour activer la mise en cache de l'objet de données, vous devez déployer le mappage sur une application, activer la mise en cache de l'objet de données logique et exécuter le mappage avec la commande `infacmd ms runmapping`. Lorsque vous exécutez le mappage, le service d'intégration de données interroge le cache d'objet de données plutôt que l'objet de données logique.

Si vous exécutez le mappage depuis l'outil Developer, la transformation Recherche interroge l'objet de données logique ligne par ligne.

Utilisez le type de cache approprié.

Utilisez les types de cache suivants pour augmenter les performances :

- Cache partagé. Vous pouvez partager le cache de recherche entre plusieurs transformations. Vous pouvez partager un cache non nommé entre des transformations dans le même mappage. Vous pouvez partager un cache nommé entre des transformations dans le même mappage ou dans des mappages différents.
- Cache persistant. Pour enregistrer et réutiliser les fichiers de cache, vous pouvez configurer la transformation pour qu'elle utilise un cache persistant. Utilisez cette fonctionnalité lorsque vous savez que la table de recherche ne change pas entre les exécutions du mappage. L'utilisation d'un cache persistant peut améliorer les performances car le Data Integration Service construit le cache mémoire à partir des fichiers cache au lieu de la base de données.

Activez les caches simultanés.

Lorsque le Data Integration Service traite des mappages contenant des transformations Recherche, il construit un cache dans la mémoire lorsqu'il traite la première ligne de données dans une transformation Recherche mise en cache. S'il y a plusieurs transformations Recherche dans un mappage, le Data Integration Service crée les caches de manière séquentielle lorsque la première ligne de données est traitée par la transformation Recherche. Cela ralentit le traitement de la transformation Recherche.

Vous pouvez activer des caches simultanés pour améliorer les performances. Lorsque le nombre de pipelines simultanés supplémentaires est défini sur un ou plus, le Data Integration Service construit des caches simultanément plutôt que de manière séquentielle. Les performances augmentent considérablement lorsque les mappages contiennent plusieurs transformations actives pouvant prendre du temps, telles que les transformations Agrégation, Jointure ou Trieur. Lorsque vous activez plusieurs pipelines simultanés, le Data Integration Service n'attend plus que des mappages actifs se terminent avant de créer le cache. D'autres transformations Recherche dans le pipeline construisent également des caches simultanément.

Optimisez les correspondances de la condition de recherche.

Lorsque la transformation Recherche correspond à des données de cache de recherche avec la condition de recherche, elle trie et ordonne les données pour déterminer la première et la dernière valeurs correspondantes. Vous pouvez configurer la transformation pour qu'elle renvoie n'importe quelle valeur correspondant à la condition de recherche. Lorsque vous configurez la transformation Recherche pour qu'elle renvoie n'importe quelle valeur correspondante, elle renvoie la première valeur correspondant à la condition de recherche. Elle n'indexe pas tous les ports comme lorsque vous la configurez pour qu'elle renvoie la première ou la dernière valeur correspondante.

Lorsque vous utilisez n'importe quelle valeur correspondante, les performances peuvent s'améliorer car la transformation n'indexe pas tous les ports, ce qui peut ralentir les performances.

Réduisez le nombre de lignes mises en cache.

Vous pouvez réduire le nombre de lignes incluses dans le cache pour augmenter les performances.

Utilisez l'option de remplacement SQL de recherche pour ajouter une clause WHERE à l'instruction SQL par défaut. Lorsque vous ajoutez une clause WHERE à une transformation recherche qui utilise un cache dynamique, utilisez une transformation filtre avant la transformation recherche pour passer des lignes dans le cache dynamique qui correspondent à la clause WHERE.

Remplacez l'instruction ORDER BY.

Par défaut, le Data Integration Service génère une instruction ORDER BY pour une recherche mise en cache. L'instruction ORDER BY contient tous les ports de recherche. Pour augmenter les performances, supprimez l'instruction ORDER BY par défaut et saisissez un remplacement ORDER BY avec moins de colonnes.

Le Data Integration Service génère toujours une instruction ORDER BY, même si vous en saisissez une dans le remplacement. Placez deux tirets « -- » après le remplacement ORDER BY pour supprimer l'instruction ORDER BY générée.

Par exemple, une transformation Recherche utilise la condition de recherche suivante :

```
ITEM_ID = IN_ITEM_ID
PRICE <= IN_PRICE
```

La transformation Recherche comporte trois ports de recherche utilisés dans le mappage, ITEM_ID, ITEM_NAME et PRICE. Lorsque vous saisissez l'instruction ORDER BY, saisissez les colonnes dans le même ordre que les ports dans la condition de recherche. Vous devez également entourer tous les mots réservés de la base de données par des guillemets.

Saisissez la requête de recherche suivante dans le remplacement SQL de recherche :

```
SELECT ITEMS_DIM.ITEM_NAME, ITEMS_DIM.PRICE, ITEMS_DIM.ITEM_ID FROM ITEMS_DIM ORDER
BY
ITEMS_DIM.ITEM_ID, ITEMS_DIM.PRICE --
```

Utilisez une machine avec plus de mémoire.

Pour augmenter les performances de mappage, exécutez le mappage sur un nœud de Data Integration Service avec une grande quantité de mémoire. Augmentez les tailles d'index et de cache de données le plus possible sans fatiguer la machine. Si le nœud du Data Integration Service a assez de mémoire, augmentez le cache pour qu'il puisse contenir toutes les données en mémoire sans paginer sur le disque.

Optimisez la condition de recherche.

Si vous incluez plusieurs conditions de recherche, placez les conditions dans l'ordre suivant pour optimiser les performances de recherche :

- Égal à (=)
- Inférieur à (<), supérieur à (>), inférieur ou égal à (<=), supérieur ou égale à (>=)
- N'est pas égal à (!=)

Filtrez les lignes de recherche.

Pour améliorer les performances, créez une condition de filtre pour réduire le nombre de lignes de recherche récupérées à partir de la source lorsque le cache de recherche est construit.

Indexez la table de recherche.

Le Data Integration Service doit interroger, trier et comparer des valeurs dans les colonnes de la condition de recherche. L'index doit inclure chaque colonne utilisée dans une condition de recherche.

Vous pouvez améliorer les performances pour les types de recherches suivants :

- Recherches mises en cache. Pour améliorer les performances, indexez les colonnes dans l'instruction ORDER BY de recherche. Le fichier journal de mappage contient l'instruction ORDER BY.
- Recherches non mises en cache. Pour améliorer les performances, indexez les colonnes dans la condition de recherche. Le Data Integration Service émet une instruction SELECT pour chaque ligne transmise à la transformation Recherche.

Optimisez plusieurs recherches.

Si un mappage contient plusieurs recherches, même avec la mise en cache activée et un tas mémoire suffisant, les recherches peuvent diminuer les performances. Réglez les transformations Recherche qui interrogent la plus grande quantité de données pour qu'elles améliorent les performances globales.

Si la table de recherche est sur la même base de données que la table source dans votre mappage, et que la mise en cache est impossible, joignez les tables dans la base de données source au lieu d'utiliser une transformation Recherche.

Optimisation de la transformation Trieur

La transformation Trieur peut ralentir les performances lorsque la mémoire RAM physique du nœud de service d'intégration de données n'a pas suffisamment de mémoire allouée pour trier les données.

Envisagez les solutions suivantes pour les goulots d'étranglement de transformation Trieur :

Allouez suffisamment de mémoire.

Pour optimiser les performances, configurez la taille de cache du trieur sur une valeur inférieure ou égale à la quantité de RAM physique disponible sur le nœud du service d'intégration de données. Allouez au moins 16 Mo de mémoire physique pour trier les données à l'aide de la transformation Trieur. La taille de cache du trieur est définie sur 16 777 216 octets par défaut. Si le service d'intégration de données ne peut pas allouer suffisamment de mémoire pour trier les données, le mappage échoue.

Si la quantité de données entrantes est supérieure à la taille de cache du trieur, le service d'intégration de données stocke temporairement les données dans le répertoire de travail de la transformation Trieur. Le service d'intégration de données requiert un espace disque au moins deux fois supérieur à la quantité de données entrantes lors du stockage des données dans le répertoire de travail.

Optimisation de la transformation SQL

À chaque fois que le service d'intégration de données traite une nouvelle requête dans un mappage, il appelle une fonction appelée SQLPrepare pour créer une procédure SQL et la transmettre à la base de données. Lorsque la requête change pour chaque ligne d'entrée, les performances peuvent être réduites.

Envisagez les solutions suivantes pour les goulots d'étranglement de transformation SQL :

N'utilisez pas d'instructions de transaction dans une requête de transformation SQL.

Lorsqu'une requête SQL contient des instructions de requête commit et rollback, le service d'intégration de données doit recréer la procédure SQL après chaque validation ou retour en arrière. Pour optimiser les performances, n'utilisez pas d'instructions de transaction dans une requête de transformation SQL.

Activez la sélection précoce ou les méthodes d'optimisation de filtre push-into avec la transformation SQL.

Pour améliorer les performances, envisagez d'activer la sélection précoce ou l'optimisation push-into avec la transformation SQL.

Optimisation de sélection précoce avec la transformation SQL

Le service d'intégration de données peut effectuer une optimisation de sélection précoce avec une transformation SQL si la condition de filtre référence uniquement les ports d'intercommunication et si la transformation SQL n'a pas d'effets secondaires.

La transformation SQL a des effets secondaires dans les cas suivants :

- La requête SQL met à jour une base de données. La requête SQL contient une instruction comme CREATE, DROP, INSERT, UPDATE, GRANT ou REVOKE.
- La transformation renvoie des lignes NULL pour les instructions SELECT qui ne renvoient aucun résultat. Les lignes peuvent contenir des valeurs de port d'intercommunication, des informations d'erreur SQL ou le champ NUMRowsAffected.

Activation de l'optimisation de sélection précoce avec la transformation SQL

Activez l'optimisation de sélection précoce dans la transformation SQL si la transformation SQL n'a pas d'effets secondaires.

1. Activez l'option **Renvoyer la sortie de base de données uniquement** dans les **Propriétés avancées** de la transformation SQL.
2. Effacez **A des effets secondaires** dans les **Propriétés avancées** de la transformation.
3. Si la transformation a un port **NumAffectedRows**, supprimez-le.

Optimisation push-into avec la transformation SQL

Avec l'optimisation push-into, le service d'intégration de données transmet la logique du filtre depuis une transformation Filtre dans le mappage vers la requête de la transformation SQL.

Utilisez les règles et instructions suivantes lorsque vous activez l'optimisation push-into avec la transformation SQL :

- La requête de la transformation SQL doit uniquement contenir des instructions SELECT.
- La requête SQL de la transformation doit être une sous-requête valide.
- La condition de filtre ne peut pas faire référence aux champs d'erreur SQL et NumRowsAffected.
- Les noms des ports de sortie doivent correspondre aux noms des colonnes dans l'instruction SQL SELECT. Lorsque vous faites référence à un port de sortie dans une condition de filtre, le service d'intégration de données transmet le nom de port d'entrée correspondant vers la requête SQL. Vous pouvez ajouter des alias au SQL si les colonnes dans la requête ne correspondent pas aux noms de ports de sortie. Par exemple : `SELECT mycolname1 AS portname1, mycolname2 AS portname2.`
- La transformation ne peut pas avoir d'effets secondaires.

Exemple d'optimisation push-into avec la transformation SQL

Une transformation SQL récupère les commandes par ID de client. Une transformation Filtre qui apparaît après la transformation SQL renvoie uniquement les lignes contenant une commande dont le montant est supérieur à 1 000.

Le service d'intégration de données transmet le filtre suivant dans une instruction SELECT de la transformation SQL :

```
orderAmount > 1000
```

Chaque instruction dans la requête SQL devient une sous-requête distincte de l'instruction SELECT qui contient le filtre.

L'instruction de requête suivante illustre l'instruction de requête d'origine comme sous-requête dans l'instruction SELECT :

```
SELECT <customerID>, <orderAmount>, ... FROM (instructions de requête d'origine) ALIAS WHERE <orderAmount> > 1000
```

Si la requête SQL contient plusieurs instructions, chacune d'elle est incluse dans une autre sous-requête. La sous-requête a la même syntaxe, y compris la clause WHERE.

Les ports *customerID* et *orderAmount* sont les noms des ports de sortie dans la transformation SQL. La sous-requête ne comprend aucun port d'intercommunication, d'erreur SQL ou de ports de statistiques SQL. Si vous transmettez plusieurs filtres dans la transformation SQL, la clause WHERE contient tous les filtres.

Activation de l'optimisation push-into avec la transformation SQL

Activez l'optimisation push-into en configurant les propriétés dans l'onglet **Propriétés avancées** de la transformation SQL.

1. Effacez **A des effets secondaires**.
2. Activez **Renvoyer la sortie de base de données uniquement**.
3. Définissez **Nombre max de lignes de sortie** sur zéro.
4. Activez l'optimisation push-into.

Cache de transformation

Lorsque vous exécutez un mappage qui utilise une transformation Agrégation, Jointure, Recherche, Rang ou Trieur, le service d'intégration de données crée des caches en mémoire pour traiter la transformation. Si le service d'intégration de données nécessite plus d'espace, il stocke les valeurs de débordement dans des fichiers de cache sur le disque.

Envisagez la solution suivante pour les goulots d'étranglement du cache de transformation :

Configurer les transformations pour qu'elles allouent suffisamment d'espace pour le stockage du cache en mémoire.

Pour améliorer leur temps de traitement, configurez les transformations Agrégation, Jointure, Recherche, Rang ou Trieur pour qu'elles allouent suffisamment d'espace pour le stockage du cache en mémoire. Lorsque vous configurez la quantité de mémoire cache sur une valeur supérieure ou égale à celle requise pour la mise en cache des données ou des index, vous augmentez les performances en réduisant la surcharge d'E/S du système. Lorsque le service d'intégration de données écrit des fichiers de cache sur le disque, le temps de traitement du service de données augmente en raison d'une surcharge d'E/S du système.

Par défaut, le service d'intégration de données configure automatiquement les spécifications de mémoire cache lors de l'exécution. Lorsque vous exécutez un mappage en mode de mise en cache automatique, vous pouvez affiner les tailles de cache pour les transformations. Analysez les statistiques de transformation dans le journal de mappage pour déterminer les tailles de cache requises pour traiter

les transformations en mémoire. Lorsque vous configurez la taille de cache pour utiliser la valeur spécifiée dans le journal de mappage, vous vous assurez qu'aucune partie de la mémoire allouée n'est gaspillée. Cependant, la taille optimale du cache dépend de la taille des données source. Consultez les journaux de mappage après des exécutions de mappage successives pour surveiller l'évolution de la taille de cache. Si vous configurez une taille de cache spécifique pour une transformation réutilisable, vérifiez que cette taille est optimale pour chaque utilisation de la transformation dans un mappage.

Élimination d'erreurs de transformation

En grand nombre, les erreurs de transformation réduisent les performances du service d'intégration de données. Avec chaque erreur de transformation, le service d'intégration de données fait une pause pour déterminer la cause de l'erreur et supprimer la ligne responsable de l'erreur du flux de données. Ensuite, le service d'intégration de données écrit généralement la ligne dans le fichier journal du mappage des journaux du service d'intégration de données.

Envisagez les solutions suivantes pour les goulots d'étranglement d'erreurs de transformation :

Consultez le fichier journal du mappage pour voir où interviennent les erreurs de transformation et évaluer les contraintes de ces transformations.

Les erreurs de transformation surviennent quand le service d'intégration de données rencontre des erreurs de conversion, des conflits de logique de mappage et toute condition définie comme une erreur, telle qu'une entrée null. Consultez le fichier journal du mappage pour voir où interviennent les erreurs de transformation. Si les erreurs sont regroupées autour de transformations particulières, évaluez les contraintes de ces transformations.

Configurez un niveau de suivi plus faible.

Si vous avez besoin d'exécuter un mappage qui génère un grand nombre d'erreurs de transformation, il est possible d'améliorer les performances en définissant un niveau de suivi plus faible. Cependant, ce n'est pas une solution à long terme pour les erreurs de transformation.

Effets secondaires de transformation

Une transformation a des effets secondaires si elle renvoie des lignes et modifie un objet ou si elle interagit avec d'autres objets ou fonctions. La transformation peut modifier une base de données, ajouter une valeur à un total, lever une exception, écrire un e-mail ou appeler d'autres fonctions avec des effets secondaires.

Le service d'intégration de données identifie les transformations qui ont des effets secondaires avant d'optimiser un mappage. Le service d'intégration de données suppose qu'une transformation a des effets secondaires lorsqu'il ne peut pas déterminer si la transformation en a.

Les transformations avec des effets secondaires limitent les possibilités d'optimisation du mappage par le service d'intégration de données. L'optimisation de sélection précoce, de nettoyage de branche, l'optimisation globale des prédicats et l'optimisation push-into altèrent les résultats du mappage si le service d'intégration de données les applique à une transformation ayant des effets secondaires. Avec l'optimisation de sélection précoce et push-into, la logique de filtre est déplacée à partir d'une transformation Filtre le plus près possible de la source. Si le filtre intervient avant la fonction à effet secondaire, les résultats du mappage changent.

Par exemple, une transformation reçoit un ID de client et renvoie des lignes contenant des informations de commande. La transformation écrit également les commandes dans un fichier. Le service d'intégration de données applique une optimisation de filtre avant d'écrire les commandes dans le fichier, le fichier reçoit moins de lignes que lorsque le filtre intervient plus tard dans le mappage. L'effet secondaire de la transformation est la fonction qui écrit les enregistrements de commande dans un fichier.

Les transformations suivantes ont des effets secondaires :

- Transformation SQL, transformation Consommateur de service Web et transformation Java, sauf si la propriété des effets secondaires est désactivée.
- Les transformations appelant une fonction `ABORT()` ou `ERROR()`, envoyant des e-mails ou appelant une procédure stockée.
- Les transformations écrivant dans des fichiers ou bases de données.
- Les transformations maintenant un nombre via un port de variable. Par exemple, `COUNT=COUNT+1`.

Les transformations SQL, Consommateur de service Web et Java ont des effets secondaires par défaut. Si vous configurez la transformation pour traiter les lignes sans effets secondaires, vous pouvez désactiver la propriété **A des effets secondaires** dans **Propriétés avancées**. Si la transformation n'a pas d'effets secondaires, vous pouvez activer l'optimisation en configurant des propriétés supplémentaires dans ces transformations.

Optimisation de la transformation Consommateur de service Web

La transformation Consommateur de service Web peut réduire les performances lorsqu'un mappage appelle le service Web plusieurs fois.

Envisagez les solutions suivantes pour les goulots d'étranglement de la transformation Consommateur de service Web :

Configurez la transformation Consommateur de service Web pour qu'elle utilise l'authentification par cookie.

Le serveur du service Web distant suit les utilisateurs du consommateur service Web grâce aux cookies. Vous pouvez augmenter les performances lorsqu'un mappage appelle un service Web plusieurs fois.

Lorsque vous projetez le port de cookie vers un message de demande du service Web, le fournisseur du service Web renvoie une valeur de cookie dans le message de réponse. Vous pouvez transmettre la valeur du cookie à une autre transformation en aval dans le mappage ou vous pouvez enregistrer la valeur du cookie dans un fichier. Lorsque vous enregistrez la valeur du cookie dans un fichier, vous pouvez configurer le cookie comme une entrée de la transformation Consommateur de service Web. Vous pouvez projeter le port de sortie du cookie vers n'importe quel groupe de sortie de la transformation Consommateur de service Web.

Activez la sélection précoce ou les méthodes d'optimisation de filtre push-into avec la transformation Consommateur de service Web.

Pour améliorer les performances, le service d'intégration de données peut appliquer la sélection précoce ou la méthode d'optimisation push-into avec la transformation Consommateur de service Web. Pour appliquer l'optimisation de sélection précoce, le service Web ne peut pas avoir d'effets secondaires et ne peut pas traiter les défauts comme des erreurs. Pour appliquer l'optimisation push-into, le service Web ne peut pas avoir d'effets secondaires, ne peut pas traiter les défauts comme des erreurs et la condition de filtre doit faire référence aux ports d'intercommunication.

Le service Web a un effet secondaire s'il effectue d'autres fonctions en plus de renvoyer une réponse à la transformation Consommateur de service Web. Le service Web a des effets secondaires s'il modifie une base de données, écrit dans un fichier, écrit des e-mails, actualise un nombre ou appelle d'autres services Web ayant des effets secondaires.

Optimisation de sélection précoce avec la transformation Consommateur de service Web

Lorsque le service d'intégration de données applique la méthode d'optimisation de sélection précoce à la transformation Consommateur de service Web, il déplace les conditions de filtre avant la transformation Consommateur de service Web dans le mappage, plus près de la source.

Activation de l'optimisation de sélection précoce avec la transformation Consommateur de service Web

Activation de l'optimisation de sélection précoce pour la transformation Consommateur de service Web si la transformation n'a pas d'effets secondaires et ne traite pas les défaillances comme des erreurs.

1. Ouvrez la vue **Propriétés avancées** de la transformation Consommateur de service Web.
2. Effacez **Traiter la défaillance comme une erreur**.
3. Effacez **A des effets secondaires**.

Optimisation push-into avec la transformation Consommateur de service Web

Vous pouvez configurer l'optimisation push-into avec la transformation Consommateur de service Web lorsque la transformation est dans une table virtuelle d'un service de données SQL. Vous pouvez configurer l'optimisation push-into avec la transformation Consommateur de service Web lorsque la transformation est dans un mappage.

Le mappage appelle le service Web pour extraire un ensemble de données ou un sous-ensemble de données selon les instructions dans la requête SQL de l'utilisateur final. La requête SQL de l'utilisateur final contient une condition de filtre facultative.

Avec l'optimisation push-into, la transformation Consommateur de service Web reçoit la valeur du filtre dans un port de filtre. Le port de filtre est un port d'entrée non connecté que vous identifiez comme port de filtre lorsque vous configurez l'optimisation push-into. Le port de filtre a une valeur par défaut qui garantit que le service Web renvoie toutes les lignes si la requête de l'utilisateur final ne contient aucun filtre. Le port de filtre n'est pas un port d'intercommunication.

Remarque: Le champ de filtre doit être inclus dans le groupe racine dans la demande de service Web.

Lorsque vous configurez un port de filtre, vous devez identifier un port de sortie dans la transformation Consommateur de service Web qui reçoit la colonne de données à partir de la réponse du service Web. Par exemple : si le port de filtre est un port d'entrée nommé EmployeeID, le port de sortie de la réponse peut être un port nommé EmployeeNum. L'outil Developer doit associer le port de filtre d'entrée à un port de sortie pour pousser la logique de filtre à partir de la table virtuelle lue vers la demande Consommateur de service Web. Les ports d'entrée pour une demande de service Web sont généralement différents des ports de sortie de la réponse du service Web.

Lorsque vous configurez un port de filtre, vous devez identifier un port de sortie dans la transformation Consommateur de service Web qui reçoit la colonne de données à partir de la réponse du service Web. Par exemple : si le port de filtre est un port d'entrée nommé EmployeeID, le port de sortie de la réponse peut être

un port nommé EmployeeNum. L'outil Developer doit associer le port de filtre d'entrée à un port de sortie pour pousser la logique de filtre à partir de la table lue vers la demande Consommateur de service Web. Les ports d'entrée pour une demande de service Web sont généralement différents des ports de sortie de la réponse du service Web.

Le champ de filtre ne peut pas être un port d'intercommunication. Lorsque vous configurez un port de filtre, la valeur par défaut du port est remplacée par la valeur de la condition de filtre, de sorte que la valeur du port de sortie/d'intercommunication est modifiée. Un filtre basé sur la sortie port d'intercommunication/de sortie renvoie des résultats inattendus.

Vous pouvez pousser plusieurs expressions de filtre vers la transformation Consommateur de service Web. Chaque condition de filtre doit être au format suivant :

```
<Field> = <Constant>
```

Les conditions de filtre doivent être liées par AND. Vous ne pouvez pas joindre des conditions avec OU.

Exemple d'optimisation push-into avec une transformation Consommateur de service Web

Un service de données SQL renvoie les commandes de tous les clients ou d'un client en particulier en fonction de la requête SQL qu'il reçoit de l'utilisateur.

Le service de données contient un objet de données logique avec les composants suivants :

Table Client

Une table de base de données Oracle contenant les informations clients.

Transformation Consommateur de service Web

Transformation qui appelle un service Web pour récupérer les dernières commandes des clients. La transformation Consommateur de service Web a des ports d'entrée pour customerID et orderNum. La transformation a des ports d'intercommunication contenant les données clients reçues de la table Client. Le port orderNum est le port de filtre et n'est pas connecté. orderNum a la valeur par défaut "**". Quand le service Web reçoit cette valeur dans la requête de service Web, il renvoie toutes les commandes.

Table virtuelle des commandes

Table virtuelle qui reçoit les données du client et de la commande depuis le service Web. L'utilisateur final interroge cette table. Les commandes contiennent une colonne client, une colonne orderID et les données du client et de la commande.

L'utilisateur final transmet la requête SQL suivante au service de données SQL :

```
SELECT * from OrdersID where customer = 23 and orderID = 56
```

Le service d'intégration de données fractionne la requête pour optimiser le mappage. Le service d'intégration de données utilise l'optimisation de sélection précoce et déplace la logique de filtre, `customer = 23`, dans la table client en lecture. Le service d'intégration de données utilise l'optimisation push-into et pousse la logique de filtre, `orderID = 56` dans le port de filtre de la transformation Consommateur de service Web. La transformation Consommateur de service Web récupère ordersID 56 pour le client 23.

Activation de l'optimisation push-into avec la transformation Consommateur de service Web

Activez l'optimisation push-into pour la transformation Consommateur de service Web si la transformation n'a pas d'effets secondaires et ne traite pas les défaillances comme des erreurs.

1. Ouvrez la vue **Propriétés avancées** de la transformation Consommateur de service Web.

2. Effacez **Traiter la défaillance comme une erreur**.
3. Effacez **A des effets secondaires**.
4. Cliquez sur le bouton **Ouvrir** dans la propriété **Optimisation push-into**.
5. Choisissez le nom du port de filtre dans la boîte de dialogue Entrée optimisée.
Vous pouvez sélectionner plusieurs ports de filtre.
6. Cliquez sur la colonne **Sortie**.
7. Pour chaque port de filtre, choisissez le port de sortie qui contient la colonne filtrée dans la réponse du service Web.
8. Entrez une valeur par défaut pour chaque port de filtre.

Remarque: Vous ne pouvez pas configurer de valeur par défaut pour un port Consommateur de service Web, excepté s'il s'agit d'un port de filtre.

CHAPITRE 5

Optimisation du mappage

Ce chapitre comprend les rubriques suivantes :

- [Présentation de l'optimisation du mappage, 44](#)
- [Méthodes d'optimisation, 45](#)
- [Optimisation du refoulement, 52](#)
- [Lecture à un seul passage, 54](#)
- [Optimisation de filtre, 55](#)
- [Optimisation de conversion du type de données, 56](#)
- [Suivi d'erreur, 56](#)

Présentation de l'optimisation du mappage

Optimisez les mappages pour permettre au service d'intégration de données de transformer et de déplacer des données efficacement. L'optimisation au niveau du mappage risque de prendre du temps, mais elle peut considérablement améliorer les performances du mappage.

Les tâches d'optimisation s'appliquent aux mappages réguliers, aux mappages de lecture et d'écriture d'objet de données logique, aux mappages de table virtuelle et aux mappages d'opérations. Concentrez-vous sur l'optimisation au niveau du mappage après avoir optimisé les cibles et les sources.

Les tâches d'optimisation s'appliquent aux mappages réguliers et aux mappages d'écriture et de lecture d'objet de données logique. Concentrez-vous sur l'optimisation au niveau du mappage après avoir optimisé les cibles et les sources.

Pour optimiser un mappage, procédez comme suit :

- Configurez le mappage avec le moins de transformations et d'expressions possible pour effectuer le plus gros volume de travail possible.
- Supprimez les liens inutiles entre les transformations pour limiter la quantité de données déplacées.
- Choisissez un niveau d'optimisation qui détermine quelles méthodes d'optimisation le service d'intégration de données peut appliquer au mappage. Lorsque le service d'intégration de données optimise un mappage, il tente de réduire la quantité de données à traiter. Par exemple, le service d'intégration de données peut utiliser l'optimisation de sélection précoce pour rapprocher un filtre de la source. Il peut utiliser la méthode d'optimisation basée sur le coût pour modifier l'ordre de traitement de jointure.
- Choisissez un type de refoulement pour activer le service d'intégration de données et déterminer s'il peut refouler la logique de transformation partielle ou complète vers la base de données source.

- Configurez la mise en cache de l'objet de données pour activer les objets de données logiques mis en cache du service d'intégration de données et accéder aux objets de données logiques prédéfinis lorsqu'il exécute un mappage. Par défaut, le service d'intégration de données extrait les données source et construit les objets de données requis lorsqu'il exécute un mappage. Les performances de mappage augmentent lorsque le service d'intégration de données peut accéder aux objets de données prédéfinis.
- Indiquez si la transformation SQL, la transformation Consommateur de service Web et la transformation Java n'ont pas d'effets secondaires lorsque vous les configurez. Certaines transformations ont des effets secondaires qui limitent l'optimisation. Par exemple, une transformation peut avoir un effet secondaire si elle écrit dans un fichier ou une base de données, ajoute une valeur à un nombre, soulève une exception ou écrit un e-mail. Dans la plupart des cas, le service d'intégration de données identifie les transformations ayant des effets secondaires qui limitent l'optimisation.

LIENS CONNEXES :

- ["Mise en cache de l'objet de données" à la page 69](#)

Méthodes d'optimisation

Le service d'intégration de données applique des méthodes d'optimisation pour réduire le nombre de lignes à traiter dans le mappage. Vous pouvez configurer le niveau d'optimisation du mappage afin de limiter les méthodes d'optimisation appliquées par le service d'intégration de données.

Le service d'intégration de données peut appliquer les méthodes d'optimisation suivantes :

- Optimisation du refoulement
- Optimisation de projection précoce
- Optimisation de sélection précoce
- Optimisation de nettoyage de branche
- Optimisation push-into
- Optimisation du prédicat
- Optimisation globale des prédicats
- Optimisation basée sur le coût
- Optimisation de jointure dataship
- Optimisation de semi-jointure

Le service d'intégration de données peut appliquer plusieurs méthodes d'optimisation à un mappage simultanément. Par exemple, le service d'intégration de données applique les méthodes d'optimisation de projection précoce, du prédicat, globale des prédicats, de nettoyage de branche et de sélection précoce ou push-into lorsque vous sélectionnez le niveau d'optimisation normal.

Niveaux d'optimisation

Le service d'intégration de données optimise les mappages en fonction du niveau d'optimisation que vous configurez. Configurez le niveau d'optimisation lorsque vous voulez que le mappage utilise un niveau d'optimisation autre que le niveau normal. Par défaut, chaque mappage utilise le niveau d'optimisation normal.

Vous pouvez choisir l'un des niveaux d'optimisation suivants :

Aucun

Le service d'intégration de données n'applique aucune optimisation.

Minimal

Le service d'intégration de données applique la méthode d'optimisation de projection précoce.

Normal

Le service d'intégration de données applique les méthodes d'optimisation de projection précoce, de sélection précoce, de nettoyage de branche, push-into, globale des prédicats et du prédicat. Normal est le niveau d'optimisation par défaut.

Complet

Le service d'intégration de données applique les méthodes d'optimisation basée sur le coût, de projection précoce, de sélection précoce, de nettoyage de branche, de prédicat, push-into, de semi-jointure et de jointure dataship.

Le service d'intégration de données applique le niveau d'optimisation normal lorsque vous exécutez un mappage à partir du menu **Exécuter** ou de l'éditeur de mappage dans l'outil Developer. Lorsque vous exécutez le mappage à partir du menu **Exécuter**, le service d'intégration de données applique le niveau d'optimisation dans la configuration du mappage. Lorsque vous exécutez le mappage à partir de la ligne de commande, le service d'intégration de données applique le niveau d'optimisation à partir des propriétés du déploiement de mappage de l'application.

Remarque: Le service d'intégration de données n'applique pas la méthode d'optimisation du refoulement avec un niveau d'optimisation. Vous pouvez configurer l'optimisation du refoulement pour un mappage dans les propriétés d'exécution du mappage.

Optimisations de filtre

L'optimisation de filtre améliore les performances en réduisant le nombre de lignes transmises via le mappage. Le service d'intégration de données peut appliquer l'optimisation de sélection précoce ou l'optimisation push-into.

Lorsque le service d'intégration de données applique une méthode d'optimisation de filtre, il déplace un filtre aussi près de la source que possible dans un mappage. Si le service d'intégration de données ne peut pas déplacer un filtre avant une transformation dans un mappage, il peut peut-être pousser la logique de filtre dans une transformation.

Méthode d'optimisation de projection précoce

Lorsque le service d'intégration de données applique la méthode d'optimisation de projection précoce, il identifie les ports inutilisés et retire les liens entre ces ports.

La méthode d'optimisation de projection précoce améliore les performances en réduisant la quantité de données que le service d'intégration de données déplace entre les transformations. Lorsque le service d'intégration de données traite un mappage, il déplace les données de tous les ports connectés dans un mappage d'une transformation vers une autre. Dans des mappages volumineux, complexes, ou dans les mappages utilisant des mapplets imbriqués, il se peut que certains ports ne fournissent pas de données à la cible. Le service d'intégration de données identifie les ports qui ne fournissent pas de données à la cible. Après avoir identifié les ports inutilisés, le service d'intégration de données retire les liens entre tous les ports inutilisés du mappage.

Le service d'intégration de données ne retire pas tous les liens. Par exemple, il ne retire pas les liens suivants :

- Liens connecté à une transformation qui a des effets secondaires.
- Liens connectés à des transformations pouvant appeler une fonction `ABORT()` ou `ERROR()`, envoyer un e-mail ou appeler une procédure stockée.

Si le service d'intégration de données détermine que tous les ports d'une transformation sont inutilisés, il retire tous les liens de la transformation excepté le lien vers le port contenant le moins de données. Le service d'intégration de données ne retire pas la transformation inutilisée du mappage.

L'outil Developer active cette méthode d'optimisation par défaut.

Méthode d'optimisation de prédicat

Lorsque le service d'intégration de données applique la méthode d'optimisation de prédicat, il examine les expressions de prédicat que génère un mappage. Il détermine s'il peut simplifier ou réécrire les expressions pour augmenter les performances du mappage.

Lorsque le service d'intégration de données exécute un mappage, il génère des requêtes sur les sources de mappage et effectue des opérations sur les résultats de requête basées sur la logique de mappage et les transformations à l'intérieur du mappage. Les requêtes et les opérations incluent souvent des expressions de prédicat. Les expressions de prédicat représentent les conditions que les données doivent respecter. Les conditions de filtre et de jointure dans les transformations Filtre et Jointure sont des exemples d'expressions de prédicat.

Avec la méthode d'optimisation de prédicat, le service d'intégration de données tente également d'appliquer des expressions de prédicat le plus tôt possible dans le mappage pour améliorer les performances de mappage.

Le service d'intégration de données déduit des relations à partir d'expressions de prédicat existantes et en crée de nouvelles. Par exemple, un mappage contient une transformation Jointure avec la condition de jointure « $A=B$ » et une transformation Filtre avec la condition de filtre « $A>5$ ». Le service d'intégration de données pourra peut-être ajouter « $B>5$ » à la condition de jointure.

Le service d'intégration de données applique la méthode d'optimisation de prédicat avec la méthode d'optimisation de sélection précoce lorsqu'il peut appliquer les deux méthodes à un mappage. Par exemple, lorsque le service d'intégration de données crée de nouvelles conditions de filtre via la méthode d'optimisation de prédicat, il tente également de les déplacer en amont dans le mappage à l'aide de la méthode de sélection précoce. L'application des deux méthodes d'optimisation améliore les performances de mappage par rapport à l'application de l'une ou l'autre méthode seule.

Le service d'intégration de données applique la méthode d'optimisation de prédicat si son application augmente les performances. Le service d'intégration de données n'applique pas cette méthode si son application modifie les résultats de mappage ou réduit les performances du mappage. Le service d'intégration de données applique cette méthode d'optimisation par défaut.

Règles et instructions d'optimisation de prédicat

Lorsque le Data Integration Service réécrit une expression de prédicat, il applique la logique mathématique à l'expression pour l'optimiser.

Le Data Integration Service peut effectuer n'importe laquelle des actions suivantes :

- Identifier des variables équivalentes dans des expressions de prédicat dans le mappage et générer des expressions simplifiées à partir des équivalences.
- Identifier des prédicats redondants dans les expressions de prédicat dans le mappage et les retirer.

- Extraire des sous-expressions de clauses disjonctives et générer plusieurs expressions simplifiées basées sur les sous-expressions.
- Normaliser une expression de prédicat.
- Appliquer des expressions de prédicat le plus tôt possible dans le mappage.

Le Data Integration Service risque de ne pas appliquer l'optimisation de prédicat à un mappage lorsque celui-ci contient des transformations avec un type de données différent entre les ports connectés.

Le Data Integration Service risque de ne pas appliquer l'optimisation de prédicat à une transformation lorsque l'une des conditions suivantes est vraie :

- La transformation contient des valeurs par défaut explicites pour les ports connectés.
- La transformation a des effets secondaires.
- La transformation n'autorise pas le déplacement des prédicats. Par exemple, une transformation qui a des effets secondaires peut présenter cette restriction.

L'outil Developer active la méthode d'optimisation de prédicat par défaut.

Méthode d'optimisation basée sur le coût

Grâce à l'optimisation basée sur le coût, le service d'intégration de données évalue un mappage, génère des mappages sémantiquement équivalents et exécute le mappage avec les meilleures performances possible. L'optimisation basée sur le coût réduit le temps d'exécution des mappages qui effectuent des opérations de jointure intérieure et full outer.

Les mappages sémantiquement équivalents sont des mappages qui effectuent des fonctions identiques et produisent les mêmes résultats. Pour générer des mappages sémantiquement équivalents, le service d'intégration de données divise le mappage d'origine en fragments. Le service d'intégration de données détermine ensuite quels fragments de mappage il peut optimiser.

Pendant l'optimisation, le service d'intégration de données peut ajouter, supprimer ou réorganiser des transformations à l'intérieur d'un fragment. Le service d'intégration de données vérifie que les fragments optimisés produisent les mêmes résultats que les fragments d'origine et forme des mappages de substitution utilisant les fragments optimisés.

Le service d'intégration de données peut également appliquer une jointure de fusion triée si les performances de cette jointure sont supérieures à celles de la jointure de boucle imbriquée. Une jointure de fusion triée utilise l'ordre de tri pour organiser deux ensembles de données avant d'effectuer la jointure. Une jointure de boucle imbriquée utilise des boucles imbriquées pour joindre deux ensembles de données. Le service d'intégration de données peut utiliser les informations de tri dans les sources ou créer une transformation Trieur si le coût entraîné par le tri des données est inférieur à celui du traitement de la jointure de boucle imbriquée.

Le service d'intégration de données génère tous ou presque tous les mappages qui sont sémantiquement équivalents au mappage d'origine. Il utilise les statistiques de base de données ou de profilage pour calculer le coût du mappage d'origine et de chaque mappage alternatif. Ensuite, il identifie le mappage qui s'exécute le plus rapidement. Le service d'intégration de données effectue un contrôle de validation sur le meilleur mappage de substitution pour garantir qu'il est valide et qu'il produit les mêmes résultats que le mappage d'origine.

Le service d'intégration de données met en cache le meilleur mappage de substitution en mémoire. Lorsque vous exécutez un mappage, le service d'intégration de données récupère le mappage de substitution et l'exécute à la place du mappage d'origine.

L'outil Developer n'active pas cette méthode par défaut.

Méthode d'optimisation de jointure dataship

La méthode d'optimisation de jointure dataship tente de placer de plus petits ensembles de données à côté de plus grands ensembles afin de réduire le temps de traitement des jointures. Le service d'intégration de données tente d'appliquer la méthode d'optimisation de jointure dataship lorsqu'une importante différence de taille est constatée entre deux tables.

Par exemple, le service d'intégration de données peut appliquer la méthode d'optimisation de jointure dataship pour joindre une table principale qui contient 10 000 lignes avec une table détaillée qui en contient 1 000 000. Pour effectuer la jointure dataship, le service d'intégration de données crée dans la base de données une table temporaire qui contient la plus grande table détaillée. Ensuite, le service d'intégration de données copie la table principale plus petite dans une table temporaire et joint les données de la table temporaire avec celles de la table détaillée plus grande. Une fois que le service d'intégration de données a effectué l'opération de jointure, la logique de la transformation Jointure est traitée dans la base de données.

Avant d'appliquer la méthode d'optimisation de jointure dataship, le service d'intégration de données effectue des analyses afin de déterminer si cette optimisation est possible et utile. Si les analyses déterminent que cette méthode est susceptible d'améliorer les performances, le service d'intégration de données l'applique au mappage. Le service d'intégration de données analyse alors de nouveau le mappage pour déterminer s'il existe d'autres opportunités d'optimisation de jointure dataship. Le cas échéant, il effectue des optimisations supplémentaires.

L'outil Developer tool n'active pas cette méthode par défaut.

Configuration requise de la jointure dataship pour des performances optimisées

La méthode d'optimisation de jointure dataship n'améliore pas toujours les performances. Les facteurs suivants affectent les performances du mappage à l'optimisation de jointure dataship :

- La source principale de la transformation Jointure doit avoir beaucoup moins de lignes que la source détaillée.
- La source détaillée doit être assez volumineuse pour justifier l'optimisation. Si la source détaillée n'est pas assez grande, le service d'intégration de données détermine qu'il est plus rapide de lire les données depuis la source principale et la source secondaire sans appliquer la méthode d'optimisation de jointure dataship.

Règles et instructions de l'optimisation de jointure dataship

Le service d'intégration de données peut appliquer une optimisation de jointure dataship à une transformation Jointure si celle-ci répond aux exigences suivantes :

- Le type de jointure doit être normal, extérieur principal ou extérieur détaillé.
- Le pipeline de détail doit provenir d'une source relationnelle.
- Si le mappage utilise des validations basées sur la cible, la portée de la transformation Jointure doit être Toutes les entrées.
- Les pipelines principal et de détail ne peuvent pas partager de transformation.
- Le mappage ne peut pas contenir de branche entre la source détaillée et la transformation Jointure.
- Le service d'intégration de données ne peut pas appliquer la méthode d'optimisation de jointure dataship si la base de données qui contient le côté de détail de la jointure est une base de données IBM DB2 qui ne prend pas en charge le codage Unicode.

Méthode d'optimisation de semi-jointure

La méthode d'optimisation de semi-jointure tente de réduire la quantité de données extraites de la source en modifiant les opérations de jointure dans le mappage.

Le service d'intégration de données applique la méthode d'optimisation de semi-jointure à une transformation Jointure lorsqu'un groupe d'entrée a beaucoup plus de lignes que l'autre, et lorsque le plus grand groupe a beaucoup de lignes sans correspondance dans le plus petit groupe en fonction de la condition de jointure. Le service d'intégration de données tente de réduire la taille de l'ensemble de données d'un opérande de jointure en lisant les lignes du plus petit groupe, en trouvant les lignes correspondantes dans le plus grand groupe, et en effectuant ensuite l'opération de jointure. Le fait de réduire la taille de l'ensemble de données améliore les performances du mappage car le service d'intégration de données ne lit plus de lignes inutiles dans la source du plus grand groupe. Le service d'intégration de données déplace la condition de jointure vers la source du plus grand groupe et ne lit que les lignes qui correspondent au plus petit groupe.

Avant d'appliquer cette méthode d'optimisation, le service d'intégration de données effectue des analyses pour déterminer si l'optimisation de semi-jointure est possible et utile. Si les analyses déterminent que cette méthode est susceptible d'améliorer les performances, le service d'intégration de données l'applique au mappage. Le service d'intégration de données analyse alors à nouveau le mappage pour déterminer s'il y a d'autres opportunités d'optimisation de semi-jointure. Le cas échéant, il effectue des optimisations supplémentaires.

L'outil Developer n'active pas cette méthode par défaut.

Exigences d'optimisation de semi-jointure pour des performances optimisées

La méthode d'optimisation de semi-jointure n'augmente pas toujours les performances. Les facteurs suivants affectent les performances du mappage avec l'optimisation de semi-jointure :

- La source principale de la transformation Jointure doit avoir beaucoup moins de lignes que la source détaillée.
- La source détaillée doit être assez volumineuse pour justifier l'optimisation. Lorsque le Data Integration Service applique l'optimisation de semi-jointure, la méthode rajoute du temps au traitement du mappage. Si la source détaillée est petite, le temps requis pour appliquer la méthode de semi-jointure peut dépasser le temps requis pour traiter toutes les lignes dans la source détaillée.
- Le Data Integration Service doit pouvoir obtenir des statistiques du nombre de lignes source pour une transformation Jointure afin de comparer de façon précise les exigences de temps de l'opération de jointure normale par rapport à l'opération de semi-jointure.

Règles et instructions d'optimisation de semi-jointure

Le Data Integration Service peut appliquer une optimisation de semi-jointure à une transformation Jointure si celle-ci répond aux exigences suivantes :

- Le type de jointure doit être normal, extérieur principal ou extérieur détaillé. La transformation Jointure ne peut pas effectuer une jointure full outer.
- Le pipeline de détail doit provenir d'une source relationnelle.
- La condition de jointure doit être une condition de tri-fusion-jointure valide. C'est-à-dire que chaque clause doit être égale à un port principal et un port de détail. S'il y a plusieurs clauses, elles doivent être jointes par AND.
- Si le mappage n'utilise pas de validations basées sur la cible, la portée de la transformation Jointure doit être Toutes les entrées.
- Les pipelines principal et de détail ne peuvent pas partager de transformation.

- Le mappage ne peut pas contenir de branche entre la source détaillée et la transformation Jointure.

Méthode d'optimisation de sélection précoce

Lorsque le service d'intégration de données applique la méthode d'optimisation de sélection précoce, il sépare, déplace ou retire les transformations Filtre dans un mappage. Il déplace les filtres vers le haut du mappage, plus près de la source.

Le service d'intégration de données peut séparer une transformation Filtre si la condition de filtre est une conjonction. Par exemple, le service d'intégration de données peut séparer la condition de filtre « A>100 AND B<50 » en deux conditions plus simples, « A>100 » et « B<50 ». Lorsque le service d'intégration de données sépare un filtre, il déplace les filtres simplifiés vers le haut du pipeline de mappage, plus près de la source. Le service d'intégration de données déplace les filtres vers le haut du pipeline de manière distincte lorsqu'il sépare le filtre.

La méthode d'optimisation de sélection précoce est activée par défaut lorsque vous choisissez le niveau d'optimisation normal ou complet dans l'outil Developer tool. Le service d'intégration de données ignore l'optimisation de sélection précoce si une transformation qui précède la transformation Filtre a des effets secondaires. Le service d'intégration de données ne peut pas déterminer si les transformations SQL, Consommateur de service Web et Java ont des effets secondaires. Vous pouvez configurer l'optimisation de sélection précoce pour ces transformations si elles n'ont pas d'effets secondaires.

Vous pouvez désactiver la sélection précoce si l'optimisation n'améliore pas les performances. Le service d'intégration de données active cette méthode d'optimisation par défaut.

Méthode d'optimisation globale des prédicats

Lorsque le service d'intégration de données utilise la méthode d'optimisation globale des prédicats, il supprime le plus tôt possible les lignes qui peuvent être filtrées dans le mappage. Cela réduit le nombre de lignes qui doivent être traitées par le mappage. La méthode d'optimisation globale des prédicats inclut les méthodes d'optimisation des prédicats et de sélection précoce.

Par exemple, un mappage contient une transformation Jointure avec la condition de jointure « A=B » et une transformation Filtre avec la condition de filtre « A>5 ». Le service d'intégration de données peut ajouter « B>5 » à la condition de jointure et rapprocher la transformation Filtre de la source.

La méthode d'optimisation globale des prédicats applique les expressions de prédicat plus efficacement que la méthode d'optimisation des prédicats. La méthode d'optimisation globale des prédicats détermine si elle peut simplifier ou réécrire les expressions pour améliorer les performances du mappage. Elle tente aussi d'appliquer des expressions de prédicat le plus tôt possible dans le mappage pour en améliorer les performances.

La méthode d'optimisation globale des prédicats déduit les filtres et les rapproche de la source lorsque le mappage contient des jointures imbriquées ou des branches comportant des filtres. Lorsque le service d'intégration de données utilise la méthode d'optimisation globale des prédicats, il sépare les filtres, les rapproche de la source ou les supprime du mappage.

Méthode d'optimisation de nettoyage de branche

Le service d'intégration de données peut appliquer la méthode d'optimisation de nettoyage de branche aux transformations qui ne fournissent aucune ligne à la cible dans un mappage.

Il est possible que le service d'intégration de données supprime une transformation Filtre si la condition de filtre renvoie la valeur FALSE pour les lignes de données. Par exemple, un mappage dispose de deux transformations Filtre qui filtrent les données provenant de deux sources relationnelles. L'une des transformations Filtre a la condition de filtre « Country=US » et l'autre « Country=Canada ». Une

transformation Union relie les deux sources relationnelles et a la condition de filtre « Country=US ». Il est possible que le service d'intégration de données supprime du mappage la transformation Filtre ayant la condition de filtre « Country=Canada ».

L'outil Developer active la méthode d'optimisation de nettoyage de branche par défaut lorsque vous choisissez le niveau d'optimisation Normal ou Complet. Vous pouvez désactiver le nettoyage de branche si l'optimisation ne permet pas d'augmenter les performances en définissant le niveau d'optimisation sur Minimal ou Aucun.

Méthode d'optimisation push-into

Avec l'optimisation push-into, le service d'intégration de données déplace la logique de transformation Filtre vers la transformation immédiatement en amont de la transformation Filtre dans le mappage. L'optimisation push-into augmente les performances en réduisant le nombre de lignes qui passent par le mappage.

Le service d'intégration de données ne déplace pas la logique de filtre dans une autre transformation si celle-ci a des effets secondaires. Le service d'intégration de données ne peut pas déterminer si les transformations SQL, Consommateur de service Web et Java ont des effets secondaires. Vous pouvez toutefois les configurer pour l'optimisation push-into.

Optimisation du refoulement

Lorsque le service d'intégration de données applique l'optimisation du refoulement, il pousse la logique de transformation vers la base de données source. Le service d'intégration de données traduit la logique de transformation en requêtes SQL et les envoie à la base de données. La base de données source exécute les requêtes SQL pour traiter les transformations.

L'optimisation du refoulement augmente les performances du mappage lorsque la base de données source peut traiter la logique de transformation plus rapidement que le service d'intégration de données. Le service d'intégration de données lit également moins de données à partir de la source.

La quantité de logique de transformation que le service d'intégration de données pousse vers la base de données de transformation source dépend de la base de données, de la logique de transformation et de la configuration du mappage. Le service d'intégration de données traite toute la logique de transformation qu'il ne peut pas pousser vers une base de données.

Lorsque vous appliquez l'optimisation du refoulement, le service d'intégration de données analyse le mappage optimisé de la source vers la cible ou jusqu'à ce qu'il atteigne une transformation en aval qu'il ne peut pas pousser vers la base de données source. Le service d'intégration de données génère et exécute une requête SELECT pour chaque source dont la logique de transformation est refoulée. Le service d'intégration de données peut également générer une requête INSERT si la cible a été poussée vers la base de données. Le service d'intégration de données lit les résultats des requêtes SQL et traite les transformations restantes du mappage.

Le service d'intégration de données applique l'optimisation du refoulement à un mappage lorsque vous sélectionnez le type de refoulement dans les propriétés d'exécution du mappage.

Vous pouvez sélectionner les types de refoulement suivants :

- Aucun. Aucun type de refoulement n'est sélectionné pour le mappage.
- Source. Le service d'intégration de données tente de refouler le plus de logique de transformation possible vers la base de données source.

- Complet. Le service d'intégration de données pousse la logique de transformation complète vers la base de données source.

Vous pouvez également créer un paramètre de chaîne pour le type de refoulement et utiliser les valeurs de paramètre suivantes :

- Aucun
- Source
- Complet

Optimisation du refoulement complet

Lorsque le service d'intégration de données applique l'optimisation du refoulement complet, il transfère toute la logique de transformation du mappage dans la base de données source. Vous pouvez configurer le refoulement complet dans les propriétés d'exécution du mappage.

L'optimisation du refoulement complet est idéale lorsque la source et la cible se trouvent dans la même base de données ou lorsque des transformations telles qu'Agrégation et Filtre sont traitées dans la base de données source et permettent de réduire la quantité de données déplacées. Par exemple, si un mappage contient une source et une cible Teradata, configurez l'optimisation du refoulement complet de façon à transférer le traitement de la logique de transformation depuis une base de données source Teradata vers une base de données cible Teradata.

Lorsque vous configurez un mappage avec une transformation Mise à jour de la stratégie pour refoulement complet, vous devez déterminer la compatibilité de refoulement pour le mappage.

Le service d'intégration de données peut refouler un mappage avec une transformation Mise à jour de la stratégie dans les cas suivants :

- La transformation cible connectée à la transformation Stratégie de mise à jour reçoit plusieurs lignes ne possédant pas la même clé.
- La transformation cible connectée à la transformation Stratégie de mise à jour reçoit plusieurs lignes possédant la même clé qui peuvent être réorganisées.

Le service d'intégration de données ne peut pas refouler un mappage avec une transformation Mise à jour de la stratégie dans le cas suivant :

- La transformation cible connectée à la transformation Stratégie de mise à jour reçoit plusieurs lignes possédant la même clé qui ne peuvent pas être réorganisées.

Vous pouvez également utiliser un paramètre de compatibilité de refoulement dans le mappage. Vous pouvez utiliser les valeurs de paramètre suivantes :

- noMultipleRowsWithSameKeyOnTarget
- reorderAllowedForMultipleRowsWithSameKey
- reorderNotAllowedForRowsWithSameKey

Le service d'intégration de données peut utiliser l'optimisation du refoulement complet pour les sources suivantes :

- Oracle
- IBM DB2
- Microsoft SQL Server
- Teradata
- Netezza

- Greenplum
- SAP HANA

Refoulement de la source

Lorsque le service d'intégration de données applique le refoulement de la source, il analyse le mappage de la source jusqu'à la cible ou jusqu'à ce qu'il accède à une transformation en aval qu'il ne peut pas forcer vers la base de données source.

Le service d'intégration de données génère et exécute une instruction SELECT basée sur la logique de transformation pour chaque transformation qu'il peut forcer vers la base de données. Ensuite, il lit les résultats de cette requête SQL et traite les transformations restantes.

Vous pouvez configurer un mappage de façon à utiliser le refoulement de la source si la source et la cible se trouvent dans des bases de données différentes. Par exemple, si un mappage contient une source Teradata et une cible Oracle, vous pouvez configurer le refoulement de la source pour forcer le traitement d'une logique de transformation vers la source Teradata.

Règles et instructions d'optimisation du refoulement

Le service d'intégration de données peut pousser la logique de transformation vers la base de données source.

Les règles et instructions suivantes s'appliquent à l'optimisation du refoulement :

- Le service d'intégration de données peut pousser la logique de transformation Recherche et Jointure vers la base de données source si les sources résident dans le même système de gestion de base de données et qu'elles utilisent des connexions identiques.
- Le service d'intégration de données ne peut pas pousser une logique de transformation vers une source contenant un type de données binaire.
- Le service d'intégration de données désactive l'optimisation du refoulement lorsque vous avez une source de données IBM DB2 et que la précision de la colonne est comprise entre 28 et 31 chiffres pour le type de données décimal.
- Le service d'intégration de données active l'optimisation du refoulement pour un service de données SQL ou un service Web par défaut. Vous ne pouvez pas désactiver l'optimisation du refoulement pour un service de données SQL ou un service Web.
- Le service d'intégration de données ne peut pas pousser une transformation Agrégation qui contient une expression avec des fonctions d'agrégation et sans agrégation dans un port qui n'est pas Grouper par.

Lecture à un seul passage

La lecture à un seul passage vous permet de remplir plusieurs cibles avec un objet de données personnalisé. Pensez à utiliser la lecture à un seul passage si vous avez plusieurs mappages utilisant les mêmes sources.

Envisagez les solutions suivantes pour les goulots d'étranglement de lecture à un seul passage :

Combinez la logique de transformation pour chaque mappage dans un mappage et utilisez un objet de données personnalisé pour chaque source.

Le Data Integration Service lit chaque source une fois puis envoie les données dans des pipelines distincts. Une ligne particulière peut être utilisée par tous les pipelines, par n'importe quelle combinaison de pipelines, ou par aucun pipeline.

Par exemple, vous avez la table source Achat, et vous utilisez cette source chaque jour pour effectuer une agrégation et un classement. Si vous placez les transformations Agrégation et Rang dans des mappages distincts, vous forcez le Data Integration Service à lire la même table source deux fois. Cependant, si vous incluez la logique d'agrégation et de classement dans un mappage avec un qualificateur source, le Data Integration Service lit la table source Achat une fois, puis envoie les données appropriées aux pipelines distincts.

Excluez les fonctions courantes des mappages.

Lorsque vous modifiez des mappages pour bénéficier de la lecture à un seul passage, vous pouvez optimiser cette fonctionnalité en excluant les fonctions courantes des mappages. Par exemple, si vous devez soustraire un pourcentage des ports Prix pour les transformations Agrégation et Rang, vous pouvez limiter le travail en soustrayant le pourcentage avant de séparer le pipeline. Vous pouvez utiliser une transformation Expression pour soustraire le pourcentage, puis séparer le mappage après la transformation.

Optimisation de filtre

Vous pouvez optimiser les mappages en filtrant à l'intérieur d'un objet de données personnalisé et en plaçant des filtres au début du mappage.

Envisagez les solutions suivantes pour filtrer les goulots d'étranglement :

Utilisez un filtre dans un objet de données personnalisé pour supprimer les lignes à la source.

Si vous filtrez les lignes à partir du mappage, vous pouvez améliorer l'efficacité en filtrant au début du flux de données. Utilisez un filtre dans un objet de données personnalisé pour supprimer les lignes à la source. L'objet de données personnalisé limite l'ensemble de lignes extrait d'une source relationnelle.

Si vous ne pouvez pas utiliser de filtre dans l'objet de données personnalisé, utilisez une transformation Filtre et déplacez-la le plus près possible de l'objet pour retirer les données inutiles au début du flux de données. La transformation Filtre limite l'ensemble de lignes envoyées à une cible.

Utilisez un filtre dans une transformation Stratégie de mise à jour si vous n'avez pas besoin de garder les lignes rejetées.

Pour améliorer les performances du mappage, vous pouvez également utiliser une transformation Filtre pour supprimer les lignes rejetées d'une transformation Stratégie de mise à jour si vous n'avez pas besoin de garder les lignes rejetées.

Évitez les expressions complexes dans les conditions de filtre.

Évitez d'utiliser des expressions complexes dans les conditions de filtre. Pour optimiser les transformations Filtre, utilisez des nombres entiers simples ou des expressions true/false dans la condition de filtre.

La transformation Filtre filtre des données à l'intérieur d'un mappage. La transformation Filtre filtre des lignes de n'importe quel type de source. L'objet de données personnalisé filtre les lignes des sources relationnelles. La transformation Filtre filtre des lignes de n'importe quel type de source.

Optimisation de conversion du type de données

Vous pouvez augmenter les performances en éliminant des conversions de type de données inutiles. Par exemple, si un mappage déplace des données d'une colonne de nombre entier vers une colonne de décimale, puis à nouveau vers une colonne de nombre entier, la conversion du type de données inutile ralentit les performances. Si possible, éliminez les conversions du type de données inutiles des mappages.

Envisagez les solutions suivantes pour les goulots d'étranglement de la conversion du type de données :
Utilisez des valeurs entières à la place d'autres types de données lorsque vous effectuez des comparaisons à l'aide des transformations Recherche et Filtre.

Par exemple, de nombreuses bases de données stockent les informations de code postal des États-Unis sous forme de que type de données Char ou Varchar. Si vous convertissez les données de code postal en type de données nombre entier, la base de données de recherche stocke le code postal 94303-1234 sous la forme 943031234. Cela permet d'augmenter la vitesse des comparaisons de recherche basées sur le code postal.

Convertissez les dates source en chaînes via des conversions port à port pour augmenter les performances du mappage.

Vous pouvez laisser les ports dans les cibles sous forme de chaînes ou les changer en ports date/heure.

Suivi d'erreur

Pour améliorer les performances, réduisez le nombre d'événements du journal générés par le service d'intégration de données lorsqu'il exécute le mappage. Améliorez les performances du mappage en actualisant le niveau d'optimisation via la configuration du mappage ou les propriétés de déploiement de mappage. Utilisez la méthode d'optimisation basée sur le coût pour optimiser les mappages.

Envisagez les solutions suivantes pour les goulots d'étranglement du suivi d'erreur :

Définissez le niveau de suivi dans les propriétés de mappage sur Concis.

Si un mappage contient un grand nombre d'erreurs de transformation et que vous n'avez pas besoin de les corriger, définissez le niveau de suivi dans les propriétés de mappage sur Concis. À ce niveau de suivi, le service d'intégration de données n'écrit pas de messages d'erreur ou d'informations de niveau de ligne pour les données rejetées.

Si vous devez déboguer le mappage et que vous définissez le niveau de suivi sur Détaillé, il se peut que vous rencontriez une importante dégradation des performances lorsque vous exécutez le mappage. N'utilisez pas le suivi détaillé lorsque vous réglez les performances. Le niveau de suivi de mappage remplace tous les niveaux de suivi spécifiques aux transformations à l'intérieur du mappage. Cela n'est pas conseillé pour répondre à long terme à un grand nombre d'erreurs de transformation.

Modifiez le niveau d'optimisation du mappage.

Si un mappage requiert un délai anormalement long pour s'exécuter, vous pouvez souhaiter modifier le niveau d'optimisation du mappage. Le niveau d'optimisation détermine quelles méthodes d'optimisation le service d'intégration de données applique au mappage à l'exécution.

Vous définissez le niveau d'optimisation pour un mappage dans la configuration du mappage ou dans les propriétés de déploiement de mappage. Le service d'intégration de données applique différents niveaux d'optimisation au mappage en fonction de la façon dont vous exécutez le mappage.

Utilisez la méthode d'optimisation basée sur le coût.

Avec l'optimisation basée sur le coût, le service d'intégration de données évalue un mappage, génère des mappages sémantiquement équivalents et exécute le mappage avec les meilleures performances. Cette méthode est plus efficace pour les mappages contenant plusieurs transformations Jointure. Elle réduit le temps d'exécution pour les mappages qui effectuent des opérations de jointure interne adjacentes, non triées.

Les mappages sémantiquement équivalents sont des mappages qui effectuent des fonctions identiques et produisent les mêmes résultats. Pour générer des mappages sémantiquement équivalents, le service d'intégration de données divise le mappage d'origine en fragments. Le service d'intégration de données détermine ensuite quels fragments de mappage il peut optimiser.

CHAPITRE 6

Optimisation du mappage partitionné

Ce chapitre comprend les rubriques suivantes :

- [Présentation de l'optimisation du mappage partitionné, 58](#)
- [Utiliser plusieurs processeurs, 59](#)
- [Augmenter la valeur de parallélisme maximal, 59](#)
- [Optimiser les fichiers plats pour le partitionnement, 60](#)
- [Optimiser les bases de données relationnelles pour le partitionnement, 61](#)
- [Optimiser les transformations pour le partitionnement, 62](#)

Présentation de l'optimisation du mappage partitionné

Si vous disposez de l'option de partitionnement, vous pouvez définir le service d'intégration de données de manière à optimiser le parallélisme lorsqu'il exécute des mappages. Lorsque vous optimisez le parallélisme, le service d'intégration de données divise dynamiquement les données sous-jacentes en partitions et traite toutes les partitions simultanément.

Si les mappages traitent des ensembles de données volumineux ou contiennent des transformations qui effectuent des calculs complexes, leur traitement peut prendre beaucoup de temps et entraîner un faible débit de données. Lorsque vous activez le partitionnement pour ces mappages, le service d'intégration de données utilise des threads supplémentaires pour traiter le mappage.

Vous pouvez optimiser les performances des mappages partitionnés en effectuant les tâches suivantes :

- Utilisez plusieurs processeurs sur les nœuds qui exécutent des mappages.
- Augmentez la valeur de parallélisme maximal du service d'intégration de données.
- Configurez les propriétés des objets de données de fichier plat.
- Configurez les bases de données relationnelles pour optimiser le partitionnement.
- Configurez les propriétés des transformations.

Utiliser plusieurs processeurs

L'augmentation du nombre de threads de traitement augmente le chargement sur les nœuds qui exécutent des mappages. Si les nœuds contiennent une large bande passante de processeur, le traitement simultané des lignes de données dans un mappage peut optimiser les performances du mappage.

Le service d'intégration de données peut utiliser plusieurs processeurs pour traiter un mappage contenant plusieurs partitions. Le nombre de processeurs que le service utilise dépend de différents facteurs tels que le nombre de points de partition, le nombre de threads créés pour chaque étape du pipeline et la quantité de ressources requise pour traiter le mappage. Un mappage simple s'exécute plus rapidement dans deux partitions, mais nécessite généralement deux fois plus de processeurs que lorsqu'il s'exécute dans une seule partition.

Augmenter la valeur de parallélisme maximal

Le parallélisme maximal détermine le nombre maximal de threads parallèles qui peuvent traiter une seule étape du pipeline. Configurez la propriété **Parallélisme maximal** du service d'intégration de données en fonction des ressources matérielles disponibles. Lorsque vous augmentez la valeur de parallélisme maximal, vous pouvez réduire le temps de traitement.

Tenez compte des instructions suivantes lorsque vous augmentez la valeur de parallélisme maximal :

Augmentez la valeur en fonction du nombre de processeurs disponibles.

Augmentez la valeur de parallélisme maximal en fonction du nombre de processeurs disponibles sur les nœuds sur lesquels des mappages s'exécutent. Lorsque vous augmentez cette valeur, le service d'intégration de données utilise plusieurs threads et plusieurs processeurs pour l'exécution du mappage. L'exécution d'un mappage simple est plus rapide dans deux partitions, mais nécessite généralement deux fois plus de processeurs que dans une seule partition.

Tenez compte du nombre total de threads de traitement.

Tenez compte du nombre total de threads de traitement lors de la définition de la valeur de parallélisme maximale. Si un mappage complexe se traduit par plusieurs points de partition supplémentaires, le service d'intégration de données peut utiliser plus de threads de traitement que le processeur ne peut gérer.

Le nombre total de threads de traitement est égal à la valeur de parallélisme maximale.

Tenez compte des autres tâches que le service d'intégration de données doit exécuter.

Si vous configurez le parallélisme maximal de telle sorte que chaque mappage utilise un grand nombre de threads, le service d'intégration de données disposera de moins de threads pour d'autres tâches.

Vous pouvez éventuellement modifier la valeur pour un mappage.

Par défaut, le parallélisme maximal de chaque mappage est défini sur Auto. Chaque mappage utilise la valeur de parallélisme maximal définie pour le service d'intégration de données.

L'outil Developer tool permet aux développeurs de modifier la valeur de parallélisme maximal dans les propriétés d'exécution du mappage afin de définir une valeur maximale pour un mappage spécifique. Lorsque le parallélisme maximal est défini sur des valeurs entières différentes pour le service d'intégration de données et le mappage, le service d'intégration de données utilise la valeur la moins élevée.

Optimiser les fichiers plats pour le partitionnement

Lorsqu'un mappage qui est activé pour le partitionnement lit depuis une source de fichier plat ou écrit dans une cible de fichier plat, le service d'intégration de données peut utiliser plusieurs threads pour lire depuis le fichier plat ou y écrire.

Optimiser les sources de fichiers plats pour le partitionnement

Pour obtenir des performances optimales lorsque plusieurs threads sont utilisés pour lire un fichier plat, configurez l'objet de données du fichier plat de sorte à optimiser le débit au lieu de conserver l'ordre des lignes.

Envisagez la solution suivante pour réduire les goulots d'étranglement dans les sources de fichiers plats partitionnés :

Configurez le partitionnement de lecture simultanée de l'objet de données du fichier plat de sorte à optimiser le débit.

Dans les propriétés avancées de l'objet de données du fichier plat, définissez la propriété **Partitionnement de lecture simultanée** pour optimiser le débit. Lorsque vous optimisez le débit, le service d'intégration de données ne conserve pas l'ordre des lignes, car il ne lit pas les lignes du fichier ou de la liste de fichiers de manière séquentielle.

Optimiser les cibles de fichiers plats pour le partitionnement

Pour obtenir des performances optimales lorsque plusieurs threads sont utilisés pour écrire dans un fichier plat, configurez les partitions pour écrire la sortie cible dans des fichiers distincts et configurez plusieurs répertoires cible.

Envisagez les solutions suivantes pour réduire les goulots d'étranglement dans les cibles de fichiers plats partitionnés :

Configurez les partitions pour écrire la sortie cible dans des fichiers distincts.

Dans les propriétés avancées de l'objet de données du fichier plat, définissez la propriété **Type de fusion** sur **Pas de fusion**. Le service d'intégration de données écrit simultanément la sortie cible dans un fichier distinct pour chaque partition. Si vous avez besoin de données cible fusionnées, le type de fusion simultanée optimise davantage les performances que le type de fusion séquentielle.

Configurez plusieurs répertoires cible.

Lorsque plusieurs threads écrivent dans le même répertoire, un goulot d'étranglement peut se produire dans le mappage en raison d'un conflit d'entrée/de sortie (E/S). Un conflit d'E/S peut se produire lorsque des threads écrivent simultanément des données dans le système de fichiers. Lorsque vous configurez plusieurs répertoires, le service d'intégration de données détermine de manière alternée le répertoire de sortie de chaque thread.

Configurez les répertoires de fichiers de sortie dans les propriétés avancées de l'objet de données du fichier plat. Utilisez la valeur du paramètre système par défaut TargetDir si l'administrateur a entré plusieurs répertoires séparés par des points-virgules pour la propriété **Répertoire cible** du service d'intégration de données dans l'outil Administrator. Vous pouvez aussi entrer une valeur différente pour configurer plusieurs répertoires de fichiers de sortie spécifiques à l'objet de données du fichier plat.

Optimiser les bases de données relationnelles pour le partitionnement

Lorsqu'un mappage activé pour le partitionnement lit ou écrit dans une base de données IBM DB2 pour LUW ou une base de données relationnelle Oracle, le service d'intégration de données peut utiliser plusieurs threads pour lire la source relationnelle ou écrire dans la cible relationnelle.

Pour optimiser les performances lorsque plusieurs threads sont utilisés pour lire ou écrire dans une base de données relationnelle DB2 pour LUW ou Oracle, vous pouvez partitionner les tables source et cible.

Remarque: Si un mappage lit ou écrit dans une base de données relationnelle autre que DB2 pour LUW ou Oracle, le service d'intégration de données utilise un thread de lecture ou un thread d'écriture.

Optimiser la base de données source pour le partitionnement

Pour obtenir des performances optimales lorsque plusieurs threads sont utilisés pour lire une base de données source DB2 pour LUW ou Oracle, vérifiez que la table source est partitionnée et configurée pour accepter les requêtes parallèles.

Pour optimiser la base de données source pour le partitionnement, procédez comme suit :

Ajoutez les partitions de base de données à la source.

Ajoutez les partitions de base de données à la source relationnelle pour augmenter la vitesse de la requête du service d'intégration de données qui lit la source. Si la source ne comporte pas de partitions de base de données, le service d'intégration de données utilise un thread pour la lire.

Activez les requêtes parallèles.

Les bases de données relationnelles peuvent avoir des options qui activent les requêtes parallèles dans la base de données. Consultez la documentation de la base de données pour en savoir plus sur ces options. Si ces options ne sont pas activées, le service d'intégration de données exécute plusieurs instructions SELECT de partition en série.

Répartissez les données dans des espaces de table différents.

Chaque base de données fournit une option permettant de répartir les données dans des espaces de table différents. Chaque espace de table peut faire référence à un système de fichiers unique, ce qui empêche les conflits d'E/S dans les partitions.

Augmentez le nombre maximal de sessions autorisées pour la base de données.

Le service d'intégration de données crée une connexion distincte à la base de données source pour chaque partition. Augmentez le nombre maximal de sessions autorisées afin que la base de données puisse gérer un plus grand nombre de connexions simultanées.

Optimiser la base de données cible pour le partitionnement

Pour obtenir des performances optimales lorsque plusieurs threads sont utilisés pour écrire dans une base de données cible DB2 pour LUW ou Oracle, vérifiez que la table cible est partitionnée et configurée pour insérer des lignes en parallèle.

Pour optimiser la base de données cible pour le partitionnement, procédez comme suit :

Ajoutez des partitions de base de données à une cible DB2 pour LUW.

Le service d'intégration de données peut utiliser plusieurs threads en vue d'écrire dans une cible DB2 pour LUW qui n'a pas de partitions de base de données. Cependant, vous pouvez améliorer les performances de chargement lorsque la cible possède des partitions de base de données. Dans ce cas,

chaque thread d'écriture se connecte au nœud DB2 pour LUW qui contient la partition de base de données. Comme les threads d'écriture se connectent à des nœuds DB2 pour LUW différents au lieu de se connecter tous au nœud maître unique, les performances augmentent.

Activez les insertions parallèles.

Les bases de données relationnelles peuvent avoir des options qui activent les insertions parallèles dans la base de données. Consultez la documentation de la base de données pour en savoir plus sur ces options. Par exemple, définissez l'option `db_writer_processes` dans une base de données Oracle et l'option `max_agents` dans une base de données DB2 pour LUW pour activer les insertions parallèles.

Répartissez les données dans des espaces de table différents.

Chaque base de données fournit une option permettant de répartir les données dans des espaces de table différents. Chaque espace de table peut faire référence à un système de fichiers unique, ce qui empêche les conflits d'E/S dans les partitions.

Augmentez le nombre maximal de sessions autorisées pour la base de données.

Le service d'intégration de données crée une connexion distincte à la base de données cible pour chaque partition. Augmentez le nombre maximal de sessions autorisées afin que la base de données puisse gérer un plus grand nombre de connexions simultanées.

Définissez les options permettant d'améliorer l'évolutivité de la base de données.

Les bases de données relationnelles peuvent avoir des options qui améliorent l'évolutivité. Par exemple, désactivez la journalisation des archives et les statistiques en fonction du temps dans une base de données Oracle pour améliorer l'évolutivité.

Optimiser les transformations pour le partitionnement

Lorsque le service d'intégration de données utilise plusieurs threads pour exécuter une transformation Agrégation, Jointure, Rang ou Trieur, le service utilise le partitionnement du cache pour répartir la taille du cache entre les threads. Pour optimiser les performances pour le partitionnement du cache, configurez plusieurs répertoires de cache.

Remarque: Une transformation Recherche ne peut utiliser qu'un seul répertoire de cache.

Envisagez la solution suivante pour réduire les goulots d'étranglement dans les transformations Agrégation, Jointure, Rang et Trieur :

Configurez plusieurs répertoires de cache.

Le partitionnement du cache crée un cache distinct pour chaque partition qui traite une transformation Agrégation, Jointure, Rang ou Trieur. Pendant le partitionnement du cache, chaque partition stocke des données différentes dans un cache distinct. Chaque cache contient les lignes requises par cette partition. Le partitionnement du cache optimise les performances de mappage, car chaque thread interroge un cache distinct en parallèle.

Si la taille du cache est inférieure à la quantité de mémoire requise pour exécuter la transformation, les threads de transformation écrivent dans le répertoire de cache pour stocker les valeurs de dépassement dans des fichiers de cache. Lorsque plusieurs threads écrivent dans le même répertoire, un goulot d'étranglement peut se produire dans le mappage en raison d'un conflit d'E/S. Un conflit d'E/S peut se produire lorsque des threads écrivent simultanément des données dans le système de fichiers. Lorsque

vous configurez plusieurs répertoires de cache, le service d'intégration de données détermine de manière alternée le répertoire de cache de chaque thread de transformation.

Dans une transformation Agrégation, Jointure ou Rang, configurez les répertoires de cache dans la propriété avancée **Répertoire de cache**. Utilisez la valeur du paramètre système par défaut CacheDir si l'administrateur a entré plusieurs répertoires séparés par des points-virgules pour la propriété **Répertoire de cache** du service d'intégration de données dans l'outil Administrator. Vous pouvez aussi entrer une valeur différente pour configurer plusieurs répertoires de cache spécifiques à la transformation.

Dans une transformation Trieur, configurez les répertoires de cache dans la propriété avancée **Répertoire de travail**. Utilisez la valeur du paramètre système par défaut TempDir si l'administrateur a entré plusieurs répertoires séparés par des points-virgules pour la propriété **Répertoires temporaires** du service d'intégration de données dans l'outil Administrator. Vous pouvez aussi entrer une valeur différente pour configurer plusieurs répertoires de cache spécifiques à la transformation.

CHAPITRE 7

Optimisation du temps d'exécution

Ce chapitre comprend les rubriques suivantes :

- [Présentation de l'optimisation de l'exécution, 64](#)
- [Optimisation du service d'application, 64](#)
- [Statistiques de surveillance, 66](#)
- [Allocation de mémoire, 68](#)
- [Mise en cache de l'objet de données, 69](#)
- [Optimisation du système, 73](#)

Présentation de l'optimisation de l'exécution

Activez les fonctionnalités de performances et réglez les propriétés du service d'intégration de données pour optimiser les performances du mappage.

Utilisez les techniques d'optimisation suivantes dans l'outil Administrator pour obtenir les meilleures performances en fonction de vos besoins :

- Optimisez les processus de service d'application.
- Configurez les statistiques de surveillance pour surveiller les goulots d'étranglement du système.
- Allouez de la mémoire pour un fonctionnement optimal du système.
- Configuration de la mise en cache de l'objet de données
- Optimisez le système pour éviter les retards système et les lenteurs d'accès au disque.

Optimisation du service d'application

Optimisez le processus de service d'application lorsque les performances sont affectées. Vous pouvez optimiser le service Analyst, le service d'intégration de données et le service de référentiel modèle. Optimisez le processus de service d'application lorsque les performances sont affectées. Vous pouvez optimiser le service d'intégration de données et le service de référentiel modèle.

Optimisation du service Analyst

Réglez le service Analyst pour optimiser les performances. Vous pouvez configurer la propriété du processus du service Analyst pour la mémoire, réduire la latence réseau et configurer les paramètres de chargement de fichier plat de l'outil Analyst pour améliorer les performances du service.

Envisagez les solutions suivantes pour les goulots d'étranglement du service Analyst :

Configurez l'outil Analyst pour qu'il se connecte à un chemin d'accès réseau pour charger des fichiers plats de plus de 10 Mo.

Les performances du processus du service Analyst peuvent diminuer si les analystes chargent des fichiers plats de plus de 10 Mo dans le répertoire d'installation Informatica sur la machine sur laquelle est exécuté l'outil Analyst. Cela peut affecter l'espace disque et les performances réseau.

Téléchargez des fichiers plats de moins de 10 Mo dans le répertoire d'installation Informatica à partir de l'outil Analyst.

Les performances du processus de service Analyst peuvent diminuer si les analystes chargent des fichiers plats de plus de 10 Mo dans le répertoire d'installation Informatica à partir de l'outil Analyst. Cela peut affecter l'espace disque et les performances réseau.

Augmentez la propriété Taille maximum du tas mémoire du processus de service Analyst.

Le processus de service Analyst peut consommer une grande quantité de mémoire lorsqu'il traite un nombre important d'utilisateurs connectés au même moment. Cela peut entraîner l'ouverture d'un grand nombre de connexions réseau entre le service Analyst et d'autres services tels que le service d'intégration de données ou le service de référentiel modèle.

Utilisez l'outil Administrator pour configurer la propriété Taille maximum du tas mémoire sur une plus grande valeur dans les Propriétés avancées du processus de service Analyst.

Exportez des spécifications de mappage de taille importante vers une table ou exportez-les vers un fichier plat et tronquez-le.

Le processus de service Analyst peut avoir un impact sur les performances lorsque les analystes exportent des spécifications de mappage de taille importante telles que des fichiers plats à partir de l'outil Analyst.

Optimisation du service d'intégration de données

Réglez le processus du Data Integration Service pour améliorer les performances du service. Vous pouvez configurer les propriétés de mémoire du processus du Data Integration Service. Vous pouvez configurer chaque service Web et service de données SQL exécuté sur un Data Integration Service pour qu'il traite des requêtes simultanées.

Envisagez les solutions suivantes pour les goulots d'étranglement du Data Integration Service :

Configurez la propriété « Taille maximum du tas mémoire » du processus du Data Integration Service.

Le Data Integration Service peut consommer une grande quantité de mémoire lors du traitement de services de données SQL et de services Web.

Utilisez l'outil Administrator pour configurer la propriété « Taille maximum du tas mémoire » sur une plus grande valeur dans les Propriétés avancées du processus du Data Integration Service.

Configurez la propriété de service Web « Durée d'activation du DTM » pour le Data Integration Service.

Le service d'intégration de données consomme des ressources système pour générer une instance DTM pour chaque demande de service Web. Configurez le service d'intégration de données de manière à utiliser une seule instance DTM pour traiter plusieurs demandes de service Web.

Utilisez l'outil Administrator pour configurer la propriété du service Web « Durée d'activation du DTM » pour le Data Integration Service.

Configurez les options d'exécution dans les propriétés du processus d'intégration de données et les propriétés du service Web et du service de données SQL pour les requêtes simultanées.

Le Data Integration Service, chaque service de données SQL et chaque service Web exécuté sur le Data Integration Service consomment des ressources système et mémoire pour chaque requête simultanée.

Pour configurer le nombre de demandes simultanées que peuvent accepter le service d'intégration de données, chaque service de données SQL et chaque service Web, configurez les propriétés du service d'intégration de données et celles du service Web.

Utilisez l'outil Administrator pour configurer les options et propriétés suivantes pour le Data Integration Service, le service Web et le service de données SQL :

- Configurez les options d'exécution du service d'intégration de données.
- Configurer la propriété Nombre maximal de connexions simultanées pour chaque service de données SQL dans les propriétés SQL du processus du service d'intégration de données.
- Configurez les propriétés Nombre maximal de requêtes en retard et Nombre maximal de requêtes simultanées pour chaque service Web dans les propriétés de configuration HTTP pour le processus du Data Integration Service.

Désactivez le niveau de trace du service Web.

Le nombre de fichiers journaux du service Web que le Data Integration Service écrit et conserve peut diminuer les performances.

Utilisez l'outil Administrator pour configurer le niveau de trace du service Web pour réduire la quantité de fichiers journaux d'exécution du service Web que le Data Integration Service stocke sur le disque.

Optimisation du Model Repository Service

Réglez le Model Repository Service pour améliorer ses performances. Vous pouvez configurer la propriété de mémoire du processus du Model Repository Service et réduire la latence réseau.

Envisagez les solutions suivantes pour les goulots d'étranglement du Model Repository Service :

Hébergez la base de données du référentiel modèle sur la même machine que le Model Repository Service.

Les performances du processus du Model Repository Service peuvent être affectées si la base de données du référentiel modèle est hébergée sur un serveur distant. Les opérations du Model Repository Service qui requièrent une communication entre le référentiel modèle et le Model Repository Service sur un réseau à latence élevée peuvent ralentir les performances du Model Repository Service.

Augmentez la propriété Taille maximum du tas mémoire du processus du Model Repository Service.

Le processus du Model Repository Service peut consommer une grande quantité de mémoire lorsqu'il traite un nombre important d'utilisateurs connectés au même moment. Cela peut entraîner l'ouverture d'un grand nombre de connexions réseau entre le Model Repository Service et d'autres services tels que le Data Integration Service ou le service Analyst.

Utilisez l'outil Administrator pour configurer la propriété Taille maximum du tas mémoire sur une valeur plus grande dans les Propriétés avancées du processus du Model Repository Service.

Statistiques de surveillance

La surveillance est une fonction de domaine effectuée par le gestionnaire de service. Le gestionnaire de service stocke la configuration de surveillance dans le référentiel modèle. Utilisez l'onglet Surveillance de

l'outil Administrator tool pour surveiller les goulots d'étranglement du système tels que le nombre total de tâches en cours d'exécution, en échec, annulées et terminées qui sont exécutées sur un service sélectionné.

Envisagez la solution suivante pour les goulots d'étranglement de statistiques de surveillance :

Configurez le domaine pour définir la surveillance.

Lorsque vous configurez la surveillance, le service d'intégration de données stocke les statistiques permanentes et les rapports de surveillance dans le référentiel modèle. Les statistiques permanentes sont les données d'historique relatives aux objets d'intégration précédemment exécutés. Les rapports de surveillance présentent les mesures principales concernant l'objet d'intégration.

Configurez les paramètres de surveillance du domaine pour spécifier le référentiel modèle qui stocke les statistiques d'exécution sur les objets déployés dans les services d'intégration de données. Les paramètres de surveillance s'appliquent à tous les services d'intégration de données figurant dans le domaine et peuvent affecter les performances des services.

Le tableau suivant décrit les paramètres de surveillance susceptibles d'affecter les performances des services :

Option	Description
Conserver le résumé des données de l'historique	Nombre de jours pendant lesquels le référentiel modèle enregistre des données moyennes. Si la purge est désactivée, le référentiel modèle enregistre les données indéfiniment. La valeur par défaut est 180. La valeur minimale est 0. La valeur maximale est 366.
Conserver les données de l'historique détaillé	Nombre de jours pendant lesquels le référentiel modèle enregistre des données par minute. Si la purge est désactivée, le référentiel modèle enregistre les données indéfiniment. La valeur par défaut est 14. La valeur minimale est 1. La valeur maximale est 14.
Purger les statistiques tous les	Intervalle de temps, en jours, après lequel le service de référentiel modèle purge des données qui sont plus anciennes que les valeurs configurées dans l'option Conserver les données historiques . La valeur par défaut est 1 jour.
Jours à	Heure du jour à laquelle le service de référentiel modèle purge les statistiques. La valeur par défaut est 01:00.
Nombre maximal d'enregistrements triables	Nombre maximal d'enregistrements pouvant être triés dans l'onglet Surveiller . Si le nombre d'enregistrements dans l'onglet Surveiller est supérieur à cette valeur, vous pouvez uniquement trier par Heure de début et Heure de fin . La valeur par défaut est 3 000.
Délai maximal pour les notifications de mise à jour	Délai maximal en secondes pendant lequel le service d'intégration de données met en mémoire tampon les statistiques avant de les stocker dans le référentiel modèle et de les afficher dans l'onglet Surveiller . Si le service d'intégration de données se ferme inopinément avant d'avoir pu stocker les statistiques dans le référentiel modèle, ces statistiques sont perdues. La valeur par défaut est 10.
Afficher les millisecondes	Inclut les millisecondes pour les champs date et heure dans l'onglet Surveillance .

Allocation de mémoire

Pour optimiser les performances de mappage, configurez les propriétés de mémoire du service d'intégration de données dans l'outil Administrator.

Le tableau suivant décrit la propriété Mémoire maximale par demande pour le module de service de mappage :

Propriété	Description
Mémoire maximale par demande	<p>Le comportement de la propriété Mémoire maximale par demande dépend des configurations suivantes du service d'intégration de données :</p> <ul style="list-style-type: none">- Le service exécute les tâches dans des processus locaux ou distants distincts ou la propriété de service Taille maximale de la mémoire est définie sur 0 (valeur par défaut). La mémoire maximale par demande correspond à la quantité maximale de mémoire, en octets, que le service d'intégration de données peut allouer à toutes les transformations qui utilisent le mode de cache automatique dans une seule demande. Le service alloue de la mémoire séparément aux transformations qui disposent d'une taille du cache spécifique. La mémoire totale utilisée par la demande peut dépasser la valeur Mémoire maximale par demande.- Le service exécute les tâches dans le processus de service d'intégration de données et la valeur de la propriété de service Taille maximale de la mémoire est supérieure à 0. La mémoire maximale par demande correspond à la quantité maximale de mémoire, en octets, que le service d'intégration de données peut allouer à une seule demande. La mémoire totale utilisée par la demande ne peut pas dépasser la valeur Mémoire maximale par demande. La valeur par défaut est 536 870 912.

Le tableau suivant décrit les options d'exécution du service d'intégration de données :

Propriété	Description
Taille maximale du pool d'exécution	<p>Nombre maximal de tâches que chaque processus du service d'intégration de données peut exécuter simultanément. Les tâches incluent les aperçus de données, les mappages, les tâches de profilage, les requêtes SQL et les demandes de services Web. Par exemple, une grille du service d'intégration de données inclut trois processus de service en cours d'exécution. Si vous définissez la valeur sur 10, chaque processus de service d'intégration de données peut exécuter jusqu'à 10 tâches simultanément. Jusqu'à 30 tâches peuvent s'exécuter simultanément sur la grille. La valeur par défaut est 10.</p> <p>Remarque: Lorsque vous augmentez la valeur de la taille du pool, le service d'intégration de données utilise davantage de ressources matérielles, telles que le processeur, la mémoire et les E/S système. Définissez cette valeur en fonction des ressources disponibles sur votre machine. Par exemple, pensez au nombre de processeurs sur la machine qui héberge le service d'intégration de données et à la quantité de mémoire disponible pour le service d'intégration de données.</p>
Taille maximale de la mémoire	<p>Quantité maximale de mémoire, en octets, que le service d'intégration de données peut allouer pour l'exécution simultanée de toutes les demandes lorsqu'il exécute les tâches dans le processus de service d'intégration de données. Lorsque le service d'intégration de données exécute les tâches dans des processus locaux ou distants distincts, il ignore cette valeur. Si vous ne voulez pas limiter la quantité de mémoire que le service d'intégration de données peut allouer, définissez cette propriété sur 0.</p> <p>Si la valeur est supérieure à 0, le service d'intégration de données utilise la propriété pour calculer la quantité maximale totale de mémoire autorisée pour l'exécution simultanée de toutes les demandes. Le service d'intégration de données calcule la taille maximale de mémoire comme suit :</p> <p>Taille maximale de la mémoire + Taille maximale du tas mémoire + Mémoire requise pour le chargement des composants de programme</p> <p>La valeur par défaut est 0.</p> <p>Remarque: si vous exécutez des profils ou des mappages de qualité des données, définissez cette propriété sur 0.</p>

Le tableau suivant décrit la propriété Taille maximale du tas mémoire pour le processus du service d'intégration de données :

Propriété	Description
Taille maximale du tas mémoire	<p>Quantité de RAM allouée à la machine virtuelle Java (JVM) qui exécute le service d'intégration de données. Utilisez cette propriété pour optimiser les performances. Ajoutez l'une des lettres suivantes à la valeur afin de spécifier les unités :</p> <ul style="list-style-type: none"> - o pour octets. - k pour kilo-octets. - m pour mégaoctets. - g pour gigaoctets. <p>La valeur par défaut est 512 mégaoctets.</p> <p>Remarque: Pensez à augmenter la taille du tas mémoire lorsque le service d'intégration de données doit traiter de grandes quantités de données.</p>

Mise en cache de l'objet de données

Le service d'intégration de données utilise la mise en cache de l'objet de données pour accéder à des objets de données logiques prédéfinis. Activez la mise en cache de l'objet de données pour améliorer les performances des mappages qui incluent des objets de données logiques. Le service d'intégration de

données utilise la mise en cache de l'objet de données pour accéder à des objets de données logiques prédéfinis et à des tables virtuelles. Activez la mise en cache de l'objet de données pour améliorer les performances des mappages, des requêtes de service de données SQL et des demandes de service Web qui incluent des objets de données logiques et des tables virtuelles.

Par défaut, le service d'intégration de données extrait les données source et construit les objets de données requis lorsqu'il exécute un mappage, une demande de service de données SQL ou une demande de service Web. Lorsque vous activez la mise en cache de l'objet de données, le service d'intégration de données peut utiliser des objets de données logiques mis en cache et des tables virtuelles.

Par défaut, le service d'intégration de données extrait les données source et construit les objets de données requis lorsqu'il exécute un mappage. Lorsque vous activez la mise en cache de l'objet de données, le service d'intégration de données peut utiliser des objets de données logiques mis en cache.

Effectuez les étapes suivantes pour configurer la mise en cache de l'objet de données pour les objets de données logiques et les tables virtuelles d'une application :

1. Configurez la connexion à la base de données du cache d'objet de données dans les propriétés de cache du service d'intégration de données.
2. Activez la mise en cache dans les propriétés des objets de données logiques ou des tables virtuelles d'une application.

Effectuez les étapes suivantes pour configurer la mise en cache de l'objet de données pour les objets de données logiques d'une application :

1. Configurez la connexion à la base de données du cache d'objet de données dans les propriétés de cache du service d'intégration de données.
2. Activez la mise en cache dans les propriétés des objets de données logiques d'une application.

Par défaut, le composant du gestionnaire de cache d'objet de données du service d'intégration de données gère les tables de mise en cache pour les objets de données logiques et les tables virtuelles dans la base de données du cache d'objet de données. Lorsque le gestionnaire de cache d'objet de données gère le cache, il insère toutes les données dans les tables de mise en cache à chaque actualisation. Si vous souhaitez mettre à jour progressivement les tables de mise en cache, vous pouvez choisir de gérer les tables vous-même à l'aide d'un client de base de données ou d'un autre outil externe. Après l'activation de la mise en cache de l'objet de données, vous pouvez configurer un objet de données logique ou une table virtuelle pour l'utilisation d'une table de mise en cache gérée par l'utilisateur.

Par défaut, le composant du gestionnaire de cache d'objet de données du service d'intégration de données gère le cache des objets de données logiques dans la base de données du cache d'objet de données. Lorsque le gestionnaire de cache d'objet de données gère le cache, il insère toutes les données dans les tables de mise en cache à chaque actualisation.

Types de données pour les tables de mise en cache

Le service d'intégration de données utilise des données à partir de tables de mise en cache lorsqu'il traite des mappages, des requêtes du service de données SQL et des demandes du service Web qui contiennent des objets mis en cache. Les types de données de table de mise en cache attendus par le service d'intégration de données peuvent être différents des types de données des objets mis en cache. Le service d'intégration de données utilise les données depuis les tables de mise en cache lorsqu'il traite des mappages qui contiennent des objets mis en cache. Les types de données de table de mise en cache attendus par le service d'intégration de données peuvent être différents des types de données des objets mis en cache.

Le gestionnaire de cache d'objet de données crée les tables de mise en cache avec les types de données attendus par le service d'intégration de données. Si vous utilisez des tables de mise en cache gérées par l'utilisateur, vérifiez que ces tables utilisent les types de données attendus par le service d'intégration de données.

Types de données de mise en cache de table virtuelle

Le tableau suivant répertorie les types de données de table de mise en cache pour les tables virtuelles :

Type de données de table virtuelle	IBM DB2	Microsoft SQL Server	Oracle
Char	Vargraphic Dbclob, pour une précision supérieure à 32 672	Nvarchar Ntext, pour une précision supérieure à 4 000	Nvarchar2 Nclob, pour une précision supérieure à 2 000
Bigint	Bigint	Bigint	Nombre
Booléen	Entier	Int	Nombre
Date	Horodatage	Datetime2	Horodatage
Double	Double	Float	Horodatage
Décimal	Décimal	Décimal	Nombre
Int	Entier	Int	Nombre
Heure	Horodatage	Datetime2	Horodatage
Horodatage	Horodatage	Datetime2	Horodatage
Varbinary	Blob	Binaire Image, pour une précision supérieure à 8 000	Brut Blob, pour une précision supérieure à 2 000
Varchar	Vargraphic Dbclob, pour une précision supérieure à 32 672	Nvarchar Ntext, pour une précision supérieure à 4 000	Nvarchar2 Nclob, pour une précision supérieure à 2 000

Types de données de mise en cache des objets de données logiques

Le tableau suivant répertorie les types de données de table de mise en cache pour les objets de données logiques :

Type de données d'objets de données logiques	DB2	Microsoft SQL Server	Oracle
Bigint	Bigint	Bigint	Nombre
Binaire	Blob	Binaire Image, pour une précision supérieure à 8 000	Brut Blob, pour une précision supérieure à 2 000
Date/heure	Horodatage	Datetime2	Horodatage
Double	Double	Float	Nombre
Décimal	Décimal	Décimal	Nombre

Type de données d'objets de données logiques	DB2	Microsoft SQL Server	Oracle
Entier	Entier	Int	Nombre
Chaîne	Vargraphic Dbclob, pour une précision supérieure à 32 672	Nvarchar Ntext, pour une précision supérieure à 4 000	Nvarchar2 Nclob, pour une précision supérieure à 2 000
Texte	Vargraphic Dbclob, pour une précision supérieure à 32 672	Nvarchar Ntext, pour une précision supérieure à 4 000	Nvarchar2 Nclob, pour une précision supérieure à 2 000

Optimisation de cache d'objet de données

Les performances du cache dépendent des performances de la base de données du cache et de la configuration des objets au sein des mappages, des services de données SQL et des services Web.

Envisagez les solutions suivantes pour augmenter les performances du cache :

Optimisez la base de données du cache.

Les performances optimales du cache dépendent de la vitesse et des performances de la base de données du cache, ainsi que de la taille du cache. Configurez la taille de cache dans la base de données du cache.

Comme le gestionnaire de cache d'objet de données doit conserver l'ancien cache pour une opération d'actualisation, le cache doit être suffisamment grand pour pouvoir stocker deux ensembles de données. Utilisez la formule suivante pour estimer la taille de cache minimum requise :

$$2 * \text{average data object size} * \text{number of data objects}$$

Par exemple, vous souhaitez mettre en cache 20 objets de données logiques et 10 tables virtuelles. Si la taille moyenne des objets est de 15 Mo, la taille de cache requise est de $2 * 15 \text{ Mo} * (20 + 10) = 900 \text{ Mo}$.

Les tables de cache sont en lecture seule. Les utilisateurs finaux ne peuvent pas mettre à jour les tables de cache avec des commandes SQL.

Définissez des clés principales et des clés étrangères pour les objets de données logiques.

Lorsque le service d'intégration de données génère des caches pour des objets de données logiques avec des clés, il crée des index. Les index peuvent augmenter les performances des requêtes sur la base de données de cache.

Mettez en cache les objets de données logiques à joindre dans un mappage.

Lorsque vous joignez des objets de données logiques mis en cache, le service d'intégration de données peut pousser la logique de transformation Jointure vers la base de données de cache même lorsque les données source sont issues de bases de données différentes.

Générez le cache d'index en fonction des colonnes dans un objet de données logique ou une table virtuelle.

Configurez le service d'intégration de données pour générer un cache d'index en fonction des colonnes dans les objets de données logiques ou les tables virtuelles. L'index peut augmenter les performances des requêtes sur la base de données de cache.

Générez le cache d'index en fonction des colonnes dans un objet de données logique.

Configurez le service d'intégration de données de sorte à générer un cache d'index en fonction des colonnes dans les objets de données logiques. L'index peut augmenter les performances des requêtes sur la base de données de cache.

Optimisation du système

Souvent, les performances diminuent parce que le mappage repose sur des connexions inefficaces ou un système de processus de service d'intégration de données surchargé. Des retards système peuvent également être entraînés par des routeurs, des commutateurs, des protocoles réseau et une utilisation par de nombreux utilisateurs. Souvent, les performances diminuent parce que le mappage repose sur des connexions inefficaces ou un système de processus de service d'intégration de données surchargé. Les retards système peuvent également être provoqués par les routeurs, commutateurs et protocoles réseau.

Des lenteurs d'accès au disque sur les bases de données source et cible, les systèmes de fichiers source et cible et les nœuds dans le domaine peuvent ralentir les performances du mappage. Demandez à l'administrateur système d'évaluer les disques durs sur les machines.

Des accès disque lents sur les bases de données et les systèmes de fichiers source et cible peuvent ralentir les performances des mappages. Demandez à l'administrateur système d'évaluer les disques durs sur les machines.

Envisagez les solutions suivantes pour les goulots d'étranglement d'optimisation du système :

Améliorez la vitesse du réseau.

Des connexions réseau lentes peuvent ralentir les performances du mappage. Demandez à l'administrateur système de déterminer si le réseau fonctionne à une vitesse optimale. Réduisez le nombre d'appels de tronçons réseau entre le processus du Data Integration Service et les bases de données.

Utilisez plusieurs processeurs.

Vous pouvez utiliser plusieurs processeurs pour exécuter plusieurs mappages en parallèle.

Réduisez la pagination.

Lorsqu'un système d'exploitation manque de mémoire physique, il commence à paginer sur le disque pour libérer de la mémoire physique. Configurez la mémoire physique pour la machine du processus du Data Integration Service pour limiter la pagination sur disque.

Utilisez la liaison de processeur.

Dans un environnement UNIX multi-processeurs, le Data Integration Service risque d'utiliser une grande quantité de ressources système. Utilisez la liaison de processeur pour contrôler l'utilisation de processeur par le processus du service d'intégration. De même, si les bases de données source et cible sont sur la même machine, utilisez la liaison de processeur pour limiter les ressources utilisées par la base de données.

CHAPITRE 8

Optimisation du service de données SQL

Ce chapitre comprend les rubriques suivantes :

- [Présentation de l'optimisation du service de données SQL, 74](#)
- [Optimisation d'outil client tiers, 75](#)
- [Niveaux d'optimisation du service de données SQL, 75](#)
- [Propriétés du service de données SQL pour la mémoire et les demandes simultanées, 79](#)
- [Cache de l'ensemble de résultats pour un service de données SQL, 80](#)
- [Conservation des données virtuelles dans les tables temporaires, 82](#)

Présentation de l'optimisation du service de données SQL

Vous pouvez optimiser les services de données SQL pour améliorer les performances lorsque les utilisateurs finaux exécutent des requêtes SQL sur ces services à l'aide d'outils client tiers. Si un service de données SQL utilise un mappage de table virtuelle, vous pouvez optimiser les sources, les transformations et le mappage.

Utilisez les techniques d'optimisation suivantes pour optimiser un service de données SQL :

- Optimisez des outils client tiers.
- Configurez le niveau d'optimisation du service de données SQL.
- Configurez les propriétés de simultanéité et de mémoire du service de données SQL pour un processus d'intégration de données.
- Configurez la mise en cache de l'objet de données pour le service de données SQL.
- Configurez la mise en cache de l'ensemble de résultats pour le service de données SQL.
- Configurez des contraintes pour les tables virtuelles dans le service de données SQL.

LIENS CONNEXES :

- [“Mise en cache de l'objet de données” à la page 69](#)

Optimisation d'outil client tiers

Les outils client tiers peuvent affecter les performances lors du traitement et de l'exécution de requêtes SQL sur un service de données SQL. Optimisez les outils client tiers que peuvent utiliser les utilisateurs finaux pour exécuter des requêtes SQL sur un service de données SQL.

Envisagez les solutions suivantes pour les goulots d'étranglement d'outil client tiers :

Envoyez les résultats de requêtes volumineuses à un fichier sur disque.

Un outil client tiers peut affecter les performances s'il affiche des résultats de requêtes volumineuses sur la fenêtre de la console.

Configurez l'outil client tiers pour désactiver le cryptage.

Un outil client tiers peut affecter les performances s'il crypte des données pendant qu'il extrait ou affiche des résultats de requête.

Configurez l'outil client tiers pour qu'il extraie un ensemble de lignes au préalable.

Un outil client tiers peut affecter les performances s'il extrait une seule ligne à la fois.

Configurez l'outil client tiers pour désactiver l'option pour qu'il lise le contenu de la table lors de son premier chargement.

Un outil client tiers peut affecter les performances si les paramètres de type de données pour les types de données BLOB et CLOB sont configurés pour lire le contenu de la table lors de son premier chargement si les types de données BLOB et CLOB ne sont pas utilisés dans la requête.

Configurez l'outil client tiers pour qu'il utilise les paramètres de format et de conversion par défaut pour la date, l'heure et l'horodatage.

Un outil client tiers peut affecter les performances si les paramètres de format et de conversion de date, d'heure et d'horodatage sont définis sur un format spécifié par l'utilisateur au lieu du format par défaut.

Désactivez l'option de débogage ou définissez-la sur « pas de débogage ».

Un outil client tiers peut affecter les performances si l'option de débogage pour exécuter la requête est définie sur « suivi ». Cela peut ralentir les performances car l'outil client tiers écrit plus de messages de journal dans le fichier de débogage pendant qu'il traite de la requête.

Niveaux d'optimisation du service de données SQL

Le service d'intégration de données optimise les services de données SQL en fonction du niveau d'optimisation que vous configurez. Configurez le niveau d'optimisation lorsque vous voulez que le service de données SQL utilise un niveau d'optimisation autre que le niveau normal. Par défaut, chaque service de données SQL utilise le niveau d'optimisation normal.

Pour comprendre la manière dont le niveau d'optimisation crée une requête optimisée pour un service de données SQL, consultez le plan de requête pour un service de données SQL. Lorsque vous affichez le plan de requête, l'outil Developer affiche une représentation graphique des requêtes optimisées en fonction du niveau d'optimisation et une représentation graphique de la requête d'origine.

Vous pouvez configurer les niveaux d'optimisation suivants :

Aucun

Le service d'intégration de données n'applique aucune optimisation.

Minimal

Le service d'intégration de données applique la méthode d'optimisation de projection précoce.

Normal

Le service d'intégration de données applique les méthodes d'optimisation de projection précoce, de sélection précoce, de nettoyage de branche, push-into, globale des prédicats et du prédicat. Normal est le niveau d'optimisation par défaut.

Complet

Le service d'intégration de données applique les méthodes d'optimisation basée sur le coût, de projection précoce, de sélection précoce, de nettoyage de branche, de prédicat, push-into, de semi-jointure et de jointure dataship.

Vous pouvez utiliser une ou plusieurs des méthodes suivantes pour configurer le niveau d'optimisation d'un service de données SQL :

- Configuration du niveau d'optimisation pour la prévisualisation des données des services de données SQL.
- Configuration du niveau d'optimisation pour les services de données SQL déployés
- Configuration du niveau d'optimisation dans la chaîne de connexion de requêtes que vous avez exécutées sur des services de données SQL déployés.

Configuration du niveau d'optimisation du service de données SQL pour la prévisualisation des données

Configurez le niveau d'optimisation que le service d'intégration de données utilise pour exécuter des requêtes SQL lorsque vous prévisualisez la sortie d'un service de données SQL.

1. Dans l'outil Developer, cliquez sur **Exécuter > Ouvrir la boîte de dialogue d'exécution**.
La boîte de dialogue **Exécuter** s'affiche.
2. Cliquez sur **Configuration de la visionneuse de données**.
3. Cliquez sur le bouton **Nouveau**.
4. Entrez un nom pour la configuration de la visionneuse de données.
5. Cliquez sur l'onglet **Avancé**.
6. Sélectionnez un niveau d'optimisation.
7. Cliquez sur **Appliquer**.
8. Cliquez sur **Fermer**

L'outil Developer crée la configuration de la visionneuse de données.

Configuration du niveau d'optimisation pour les services de données SQL déployés

Configurez le niveau d'optimisation que les services d'intégration de données utilisent pour exécuter des requêtes SQL sur un service de données SQL déployé. Vous pouvez choisir de remplacer le niveau

d'optimisation d'une requête unique en configurant le niveau d'optimisation dans la connexion de service de données SQL.

1. Dans l'outil Administrator, sélectionnez un service d'intégration de données.
2. Cliquez sur la vue **Applications**.
3. Développez l'application qui contient le service de données SQL pour lequel vous voulez configurer le niveau d'optimisation.
4. Sélectionnez le service de données SQL et modifiez la propriété suivante :

Propriété	Description
Niveau d'optimisation	Niveau d'optimisation que le service d'intégration de données applique à l'objet. Entrez la valeur numérique associée au niveau d'optimisation que vous voulez configurer. Vous pouvez entrer l'une des valeurs numériques suivantes : <ul style="list-style-type: none"> - 0. Le service d'intégration de données n'applique pas d'optimisation. - 1. Le service d'intégration de données applique la méthode d'optimisation de projection précoce. - 2. Le service d'intégration de données applique les méthodes d'optimisation de projection précoce, de sélection précoce, push-into et de prédicat. - 3. Le service d'intégration de données applique les méthodes d'optimisation de projection précoce, de sélection précoce, push-into, de prédicat et de semi-jointure basées sur les coûts.

5. Pour remplacer le niveau d'optimisation que les services d'intégration de données utilisent pour exécuter une requête, joignez l'entrée suivante à la chaîne de connexion JDBC URL ou ODBC :
`SQLDataServiceOptions.optimizeLevel= <numeric_optimizer_level>.`

Plan de requête du service de données SQL

Lorsque vous affichez le plan de requête d'un service de données SQL, vous affichez la représentation graphique de la requête d'origine et la représentation graphique de la requête optimisée. La représentation graphique décrit la manière dont les processus du service d'intégration de données traitent la requête. Il inclut les transformations et l'ordre dans lequel le service d'intégration de données traite chaque transformation.

L'outil Developer utilise le niveau d'optimisation que vous définissez dans l'outil Developer pour générer la requête optimisée. La requête optimisée affiche la requête telle qu'elle est exécutée par le service d'intégration de données.

Si par exemple, vous voulez demander la table virtuelle CUSTOMERS dans un service de données SQL. Dans la vue **Visionneuse de données**, vous pouvez sélectionner les paramètres de configuration par défaut de la visionneuse de données, ce qui définit le niveau d'optimisation de la requête « normal ».

Saisissez la requête suivante dans la vue **Visionneuse de données** :

```
select * from CUSTOMERS where CUSTOMER_ID > 150000 order by LAST_NAME
```

Lorsque vous affichez le plan de requête SQL, l'outil Developer affiche la représentation graphique de la requête :



La vue non optimisée affiche la requête que vous saisissez. L'outil Developer affiche la clause WHERE comme une transformation Filtre et la clause ORDER BY comme une transformation Trieur. L'outil Developer utilise la transformation d'intercommunication Expression pour renommer les ports.

Lorsque vous affichez la requête optimisée, l'outil Developer affiche la représentation graphique de la requête :



La vue optimisée affiche la requête que le service d'intégration de données exécute. Le service d'intégration de données force la condition de filtre vers l'objet de données sources car le niveau d'optimisation est normal. Forcer la condition de filtre augmente les performances des requêtes car cela réduit le nombre de lignes que le service d'intégration de données lit dans l'objet de données sources. Tout comme pour la requête non optimisée, l'outil Developer affiche la clause ORDER BY comme une transformation Trieur. Il utilise les transformations d'intercommunication Expression pour forcer les types de données que vous spécifiez dans les transformations logiques.

Consultation d'un plan de requête SQL

Affichez le plan de requête SQL pour consulter une représentation similaire à un mappage de la requête que vous entrez lorsque vous prévisualisez les données de la table virtuelle.

1. Ouvrez un service de données SQL qui contient au moins une table virtuelle.
2. Cliquez sur la vue **Visionneuse de données**.
3. Entrez une requête SQL dans la fenêtre **Entrée**.
4. Vous pouvez également sélectionner une configuration de la visionneuse de données qui contient le niveau d'optimisation que vous voulez appliquer à la requête.
5. Cliquez sur **Afficher le plan de requête**.

L'outil Developer affiche le plan de requête SQL pour la requête que vous avez entrée dans l'onglet **Non optimisé**.

6. Pour afficher la requête optimisée, cliquez sur l'onglet **Optimisé**.

L'outil Developer affiche le plan de requête SQL optimisé.

Propriétés du service de données SQL pour la mémoire et les demandes simultanées

Pour optimiser les performances du service de données SQL, configurez les propriétés de simultanéité et de mémoire du service d'intégration de données dans l'outil Administrator.

Le tableau suivant décrit la propriété Mémoire maximale par demande pour le module de service SQL :

Propriété	Description
Mémoire maximale par demande	<p>Le comportement de la propriété Mémoire maximale par demande dépend des configurations suivantes du service d'intégration de données :</p> <ul style="list-style-type: none">- Le service exécute les tâches dans des processus locaux ou distants distincts ou la propriété de service Taille maximale de la mémoire est définie sur 0 (valeur par défaut). La mémoire maximale par demande correspond à la quantité maximale de mémoire, en octets, que le service d'intégration de données peut allouer à toutes les transformations qui utilisent le mode de cache automatique dans une seule demande. Le service alloue de la mémoire séparément aux transformations qui disposent d'une taille du cache spécifique. La mémoire totale utilisée par la demande peut dépasser la valeur Mémoire maximale par demande.- Le service exécute les tâches dans le processus de service d'intégration de données et la valeur de la propriété de service Taille maximale de la mémoire est supérieure à 0. La mémoire maximale par demande correspond à la quantité maximale de mémoire, en octets, que le service d'intégration de données peut allouer à une seule demande. La mémoire totale utilisée par la demande ne peut pas dépasser la valeur Mémoire maximale par demande. La valeur par défaut est 50 000 000.

Le tableau suivant décrit la propriété Taille maximale du tas mémoire pour le processus du service d'intégration de données :

Propriété	Description
Taille maximale du tas mémoire	<p>Quantité de RAM allouée à la machine virtuelle Java (JVM) qui exécute le service d'intégration de données. Utilisez cette propriété pour optimiser les performances. Ajoutez l'une des lettres suivantes à la valeur afin de spécifier les unités :</p> <ul style="list-style-type: none">- o pour octets.- k pour kilo-octets.- m pour mégaoctets.- g pour gigaoctets. <p>La valeur par défaut est 512 mégaoctets. Remarque: Pensez à augmenter la taille du tas mémoire lorsque le service d'intégration de données doit traiter de grandes quantités de données.</p>

Le tableau suivant décrit les propriétés SQL du processus du service d'intégration de données :

Propriété	Description
Nombre maximal de connexions simultanées	Limite le nombre de connexions à la base de données que le service d'intégration de données peut effectuer pour les services de données SQL. La valeur par défaut est 100.

Le tableau suivant décrit les options d'exécution du service d'intégration de données :

Propriété	Description
Taille maximale du pool d'exécution	<p>Nombre maximal de tâches que chaque processus du service d'intégration de données peut exécuter simultanément. Les tâches incluent les aperçus de données, les mappages, les tâches de profilage, les requêtes SQL et les demandes de services Web. Par exemple, une grille du service d'intégration de données inclut trois processus de service en cours d'exécution. Si vous définissez la valeur sur 10, chaque processus de service d'intégration de données peut exécuter jusqu'à 10 tâches simultanément. Jusqu'à 30 tâches peuvent s'exécuter simultanément sur la grille. La valeur par défaut est 10.</p> <p>Remarque: Lorsque vous augmentez la valeur de la taille du pool, le service d'intégration de données utilise davantage de ressources matérielles, telles que le processeur, la mémoire et les E/S système. Définissez cette valeur en fonction des ressources disponibles sur votre machine. Par exemple, pensez au nombre de processeurs sur la machine qui héberge le service d'intégration de données et à la quantité de mémoire disponible pour le service d'intégration de données.</p>
Taille maximale de la mémoire	<p>Quantité maximale de mémoire, en octets, que le service d'intégration de données peut allouer pour l'exécution simultanée de toutes les demandes lorsqu'il exécute les tâches dans le processus de service d'intégration de données. Lorsque le service d'intégration de données exécute les tâches dans des processus locaux ou distants distincts, il ignore cette valeur. Si vous ne voulez pas limiter la quantité de mémoire que le service d'intégration de données peut allouer, définissez cette propriété sur 0.</p> <p>Si la valeur est supérieure à 0, le service d'intégration de données utilise la propriété pour calculer la quantité maximale totale de mémoire autorisée pour l'exécution simultanée de toutes les demandes. Le service d'intégration de données calcule la taille maximale de mémoire comme suit :</p> <p>Taille maximale de la mémoire + Taille maximale du tas mémoire + Mémoire requise pour le chargement des composants de programme</p> <p>La valeur par défaut est 0.</p> <p>Remarque: si vous exécutez des profils ou des mappages de qualité des données, définissez cette propriété sur 0.</p>

Cache de l'ensemble de résultats pour un service de données SQL

Lorsque vous configurez le cache de l'ensemble de résultats, le service d'intégration de données met en cache les résultats du processus DTM associé à chaque requête du service de données SQL et chaque requête du service Web. Le service d'intégration de données met en cache les résultats de la période d'expiration que vous définissez. Quand un client effectue la même requête avant l'expiration du cache, le service d'intégration de données renvoie les résultats mis en cache.

Envisagez la solution suivante pour les goulots d'étranglement de cache de l'ensemble de résultats :

Configurez le cache de l'ensemble de résultats pour un service de données SQL.

Le cache de l'ensemble des résultats permet au service d'intégration de données d'utiliser les résultats mis en cache pour les requêtes du service de données SQL. Les utilisateurs qui exécutent des requêtes identiques dans un laps de temps court souhaitent peut-être utiliser le cache de l'ensemble de résultats pour diminuer le temps d'exécution de requêtes identiques.

Lorsque vous activez le service d'intégration de données pour qu'il utilise les résultats mis en cache, les performances du service de données augmentent. Cependant, pour améliorer davantage le temps de traitement du service de données pour des requêtes identiques, allouez suffisamment d'espace pour stocker le cache en mémoire. Lorsque vous configurez la quantité de mémoire cache pour qu'elle soit

supérieure ou égale à ce qui est requis pour mettre en cache les résultats, vous augmentez les performances en réduisant la surcharge d'E/S du système. Lorsque le service d'intégration de données écrit des fichiers cache sur le disque, le temps de traitement du service de données augmente en raison d'une surcharge d'E/S du système.

Propriétés de cache de l'ensemble de résultats de service de données SQL

Pour augmenter les performances, vous pouvez configurer la propriété de cache de l'ensemble de résultats pour un Data Integration Service. Vous pouvez également configurer le nombre de millisecondes pendant lesquelles le cache de l'ensemble de résultats est disponible pour être utilisé par un service de données SQL.

Le tableau suivant décrit les propriétés du cache de l'ensemble de résultats pour le Data Integration Service :

Propriété	Description
Préfixe de nom de fichier	Préfixe des noms de tous les fichiers de cache de l'ensemble de résultats stockés sur le disque. La valeur par défaut est RSCACHE.
Activer le cryptage	Indique si les fichiers de cache de l'ensemble de résultats sont cryptés via le cryptage AES 128 bits. Les valeurs valides sont True ou False. La valeur par défaut est True.

Le tableau suivant décrit la propriété qui configure le nombre de millisecondes pendant lesquelles le cache de l'ensemble de résultats est disponible pour le service de données SQL :

Propriété	Description
Période d'expiration du cache de l'ensemble de résultats	Nombre de millisecondes pendant lequel le cache de l'ensemble de résultats est utilisable. Si la valeur définie est -1, le cache n'expire jamais. Si la valeur définie est 0, la mise en cache de l'ensemble de résultats est désactivée. Les modifications de la période d'expiration ne s'appliquent pas aux caches existants. Pour que tous les caches utilisent la même période d'expiration, purgez le cache de l'ensemble de résultats après avoir modifié la période d'expiration. La valeur par défaut est 0.

Activation de la mise en cache de l'ensemble des résultats pour un service de données SQL

Pour utiliser des résultats mis en cache pour des requêtes de service de données SQL identiques, configurez le Data Integration Service pour qu'il utilise la mise en cache de l'ensemble des résultats.

1. Dans l'outil Administrator, sélectionnez un Data Integration Service.
2. Cliquez sur la vue **Processus** pour configurer les propriétés de mise en cache de l'ensemble des résultats.
3. Cliquez sur la vue **Application** puis cliquez sur le service de données SQL pour configurer la propriété Expiration du cache de l'ensemble des résultats.

Conservation des données virtuelles dans les tables temporaires

Une table temporaire est une table dans une base de données relationnelle qui stocke des données intermédiaires, ou « temporaires ». Les requêtes complexes nécessitent souvent une grande quantité d'espace pour stocker les données intermédiaires, telles que les informations des jointures. Si des tables temporaires existent, les outils de veille stratégique peuvent récupérer ces données dans la table temporaire plutôt que dans le service de données SQL. Les performances en sont améliorées.

Les tables temporaires offrent également une meilleure sécurité, pour deux raisons : d'abord, seul l'utilisateur de la session active peut accéder aux tables. Ensuite, les tables sont conservées tout au long de la session, avant d'être annulées par la base de données lorsque la connexion se ferme.

Implémentation des tables temporaires

Vous pouvez utiliser des tables temporaires pour améliorer les performances des requêtes volumineuses et complexes. Les tables temporaires améliorent les performances, car les requêtes dirigées vers ces tables dans une base de données relationnelle sont plus rapides que des requêtes répétées au service de données SQL portant sur le même ensemble de données.

L'implémentation de tables temporaires pour améliorer les performances nécessite l'intervention de l'administrateur Informatica et d'un développeur d'outils de veille stratégique.

Tout d'abord, l'administrateur Informatica crée une connexion de base de données relationnelle et configure le service d'intégration de données pour qu'il utilise cette connexion.

Ensuite, le développeur d'un outil de veille stratégique (par exemple, IBM Cognos ou SAP Business Objects) crée une connexion entre l'outil de veille stratégique et le service de données SQL Informatica. La connexion utilise le pilote JDBC ou ODBC Informatica.

Lorsque ces connexions sont actives, l'outil de veille stratégique peut créer et utiliser des tables temporaires pour traiter une grande quantité de données intermédiaires.

CHAPITRE 9

Optimisation du service Web

Ce chapitre comprend les rubriques suivantes :

- [Présentation de l'optimisation du service Web, 83](#)
- [Optimisation des requêtes HTTP, 84](#)
- [Compression de messages du service Web, 84](#)
- [Niveau d'optimisation du service Web , 84](#)
- [Propriétés des services Web pour la mémoire et les demandes simultanées , 86](#)
- [Propriété du service Web pour configurer une instance DTM active, 88](#)
- [Mise en cache de l'ensemble de résultats du service Web, 89](#)
- [Gestion du journal du service Web, 90](#)

Présentation de l'optimisation du service Web

Vous pouvez optimiser les services Web pour améliorer les performances lorsque le service d'intégration de données exécute des requêtes de service Web. Réglez le service d'intégration de données pour qu'il gère la mémoire et traite des requêtes de service Web simultanées. Pour améliorer les performances du service Web, utilisez la compression de message de service Web, optimisez les requêtes HTTP, configurez le cache d'objet de données et de l'ensemble de résultats et configurez les niveaux du journal d'erreur.

Utilisez les techniques d'optimisation suivantes pour optimiser un service Web :

- Optimisez les requêtes HTTP.
- Comprimez les messages du service Web.
- Configuration du niveau d'optimisation de service Web.
- Configurez les propriétés de simultanéité et de mémoire des services Web pour un processus d'intégration de données.
- Configurez le service d'intégration de données pour qu'il garde un processus DTM actif afin de pouvoir traiter plusieurs requêtes de service Web.
- Configuration de la mise en cache de l'objet de données pour le service Web.
- Configuration de la mise en cache de l'ensemble de résultats pour les services Web.
- Configurez les niveaux du journal d'erreur d'exécution des services Web.

LIENS CONNEXES :

- [“Mise en cache de l'objet de données” à la page 69](#)

Optimisation des requêtes HTTP

Optimisez les requêtes HTTP pour réduire le nombre de requêtes du serveur Web.

Envisagez les solutions suivantes pour les goulots d'étranglement de requête HTTP :

Réduisez le dépassement de délai de l'interface de connexion HTTP pour le client de service Web.

Le dépassement de délai de l'interface de connexion définit le temps que le client attend avant que le délai de la requête HTTP soit dépassé. Le client de service Web peut se bloquer si la valeur du dépassement de délai de l'interface de connexion est trop grande.

Compression de messages du service Web

Vous pouvez optimiser les performances du service Web en compressant les messages Web volumineux qui sont transmis aux fournisseurs ou sont en provenance de ceux-ci.

Envisagez la solution suivante pour les goulots d'étranglement de messages du service Web :

Activez la compression de messages SOAP pour un client de service Web.

La compression de messages SOAP permet au service Web de compresser le service Web pour recevoir des messages compressés de client du service Web. Le service Web peut accepter un message SOAP avec la compression GZip de la part d'un client de service Web.

Lorsque le service d'intégration de données reçoit la réponse du service Web, il vérifie l'en-tête HTTP de codage de contenu dans le message SOAP et décode celui-ci.

Niveau d'optimisation du service Web

Le service d'intégration de données optimise les services Web en fonction du niveau d'optimisation que vous configurez. Configurez le niveau d'optimisation lorsque vous voulez que le service Web utilise un niveau d'optimisation autre que le niveau normal. Par défaut, chaque service Web utilise le niveau d'optimisation normal.

Vous pouvez choisir l'un des niveaux d'optimisation suivants :

Aucun

Le service d'intégration de données n'applique aucune optimisation.

Minimal

Le service d'intégration de données applique la méthode d'optimisation de projection précoce.

Normal

Le service d'intégration de données applique les méthodes d'optimisation de projection précoce, de sélection précoce, de nettoyage de branche, push-into, globale des prédicats et du prédicat. Normal est le niveau d'optimisation par défaut.

Complet

Le service d'intégration de données applique les méthodes d'optimisation basée sur le coût, de projection précoce, de sélection précoce, de nettoyage de branche, de prédicat, push-into, de semi-jointure et de jointure dataship.

Vous pouvez utiliser une ou plusieurs des méthodes suivantes pour configurer le niveau d'optimisation d'un service Web :

- Configurez le niveau d'optimisation de la prévisualisation des données d'un service Web avant de le déployer sur un service d'intégration de données.
- Configurez le niveau d'optimisation des services Web déployés et exécutés sur un service d'intégration de données spécifique.
- Configurez le niveau d'optimisation dans l'en-tête de la demande de service Web pour un service Web déployé.

Configuration du niveau d'optimisation du service Web pour la prévisualisation des données

Configurez le niveau d'optimisation que les services d'intégration de données utilisent pour prévisualiser la sortie d'un service Web.

1. Dans l'outil Developer, cliquez sur **Exécuter** > **Ouvrir la boîte de dialogue d'exécution**.
La boîte de dialogue **Exécuter** s'affiche.
2. Cliquez sur **Configuration de service Web**.
3. Cliquez sur le bouton **Nouveau**.
4. Entrez un nom pour la configuration de service Web.
5. Cliquez sur l'onglet **Avancé**.
6. Sélectionnez un niveau d'optimisation.
7. Cliquez sur **Appliquer**.
8. Cliquez sur **Fermer**
L'outil Developer crée la configuration de service Web.

Lorsque vous exécutez la visionneuse de données pour prévisualiser la sortie d'un mappage d'opération, sélectionnez la configuration du Service Web qui comprend le niveau d'optimisation que vous voulez utiliser.

Configuration du niveau d'optimisation pour les services Web déployés

Configurez le niveau d'optimisation que les services d'intégration de données utilisent pour exécuter un service Web déployé. Vous pouvez choisir de remplacer le niveau d'optimisation d'une requête unique en configurant le niveau d'optimisation dans l'en-tête HTTP de la demande SOAP du service Web.

1. Dans l'outil Administrator, sélectionnez un service d'intégration de données.
2. Cliquez sur la vue **Applications**.

3. Développez l'application qui contient le service Web pour lequel vous voulez configurer le niveau d'optimisation.
4. Sélectionnez le service Web et modifiez la propriété suivante :

Propriété	Description
Niveau d'optimisation	<p>Niveau d'optimisation que le service d'intégration de données applique à l'objet. Entrez la valeur numérique associée au niveau d'optimisation que vous voulez configurer. Vous pouvez entrer l'une des valeurs numériques suivantes :</p> <ul style="list-style-type: none"> - 0. Le service d'intégration de données n'applique pas d'optimisation. - 1. Le service d'intégration de données applique la méthode d'optimisation de projection précoce. - 2. Le service d'intégration de données applique les méthodes d'optimisation de projection précoce, de sélection précoce, push-into et de prédicat. - 3. Le service d'intégration de données applique les méthodes d'optimisation de projection précoce, de sélection précoce, push-into, de prédicat et de semi-jointure basées sur les coûts.

5. Pour remplacer le niveau d'optimisation du service Web pour une demande de service Web, incluez l'entrée suivante dans l'en-tête HTTP de la demande SOAP : `WebServiceOptions.optimizeLevel=<numeric_optimizer_level>`.

Propriétés des services Web pour la mémoire et les demandes simultanées

Pour optimiser les performances du service Web, configurez les propriétés de simultanéité et de mémoire du service d'intégration de données et chaque service Web dans l'outil Administrator.

Le tableau suivant décrit la propriété Mémoire maximale par demande pour le module de service Web :

Propriété	Description
Mémoire maximale par demande	<p>Le comportement de la propriété Mémoire maximale par demande dépend des configurations suivantes du service d'intégration de données :</p> <ul style="list-style-type: none"> - Le service exécute les tâches dans des processus locaux ou distants distincts ou la propriété de service Taille maximale de la mémoire est définie sur 0 (valeur par défaut). <p>La mémoire maximale par demande correspond à la quantité maximale de mémoire, en octets, que le service d'intégration de données peut allouer à toutes les transformations qui utilisent le mode de cache automatique dans une seule demande. Le service alloue de la mémoire séparément aux transformations qui disposent d'une taille du cache spécifique. La mémoire totale utilisée par la demande peut dépasser la valeur Mémoire maximale par demande.</p> <ul style="list-style-type: none"> - Le service exécute les tâches dans le processus de service d'intégration de données et la valeur de la propriété de service Taille maximale de la mémoire est supérieure à 0. <p>La mémoire maximale par demande correspond à la quantité maximale de mémoire, en octets, que le service d'intégration de données peut allouer à une seule demande. La mémoire totale utilisée par la demande ne peut pas dépasser la valeur Mémoire maximale par demande.</p> <p>La valeur par défaut est 50 000 000.</p>

Le tableau suivant décrit les options d'exécution du service d'intégration de données :

Propriété	Description
Taille maximale du pool d'exécution	<p>Nombre maximal de tâches que chaque processus du service d'intégration de données peut exécuter simultanément. Les tâches incluent les aperçus de données, les mappages, les tâches de profilage, les requêtes SQL et les demandes de services Web. Par exemple, une grille du service d'intégration de données inclut trois processus de service en cours d'exécution. Si vous définissez la valeur sur 10, chaque processus de service d'intégration de données peut exécuter jusqu'à 10 tâches simultanément. Jusqu'à 30 tâches peuvent s'exécuter simultanément sur la grille. La valeur par défaut est 10.</p> <p>Remarque: Lorsque vous augmentez la valeur de la taille du pool, le service d'intégration de données utilise davantage de ressources matérielles, telles que le processeur, la mémoire et les E/S système. Définissez cette valeur en fonction des ressources disponibles sur votre machine. Par exemple, pensez au nombre de processeurs sur la machine qui héberge le service d'intégration de données et à la quantité de mémoire disponible pour le service d'intégration de données.</p>
Taille maximale de la mémoire	<p>Quantité maximale de mémoire, en octets, que le service d'intégration de données peut allouer pour l'exécution simultanée de toutes les demandes lorsqu'il exécute les tâches dans le processus de service d'intégration de données. Lorsque le service d'intégration de données exécute les tâches dans des processus locaux ou distants distincts, il ignore cette valeur. Si vous ne voulez pas limiter la quantité de mémoire que le service d'intégration de données peut allouer, définissez cette propriété sur 0.</p> <p>Si la valeur est supérieure à 0, le service d'intégration de données utilise la propriété pour calculer la quantité maximale totale de mémoire autorisée pour l'exécution simultanée de toutes les demandes. Le service d'intégration de données calcule la taille maximale de mémoire comme suit :</p> <p>Taille maximale de la mémoire + Taille maximale du tas mémoire + Mémoire requise pour le chargement des composants de programme</p> <p>La valeur par défaut est 0.</p> <p>Remarque: si vous exécutez des profils ou des mappages de qualité des données, définissez cette propriété sur 0.</p>

Le tableau suivant décrit les propriétés de la configuration HTTP pour le processus de service d'intégration de données :

Propriété	Description
Nombre maximal de demandes en retard	<p>Nombre maximal de connexions HTTP ou HTTPS pouvant patienter dans une file d'attente pour ce processus de service d'intégration de données. La valeur par défaut est 100.</p>
Nombre maximal de demandes simultanées	<p>Nombre maximal de connexions HTTP ou HTTPS pouvant être établies à ce processus de service d'intégration de données. La valeur par défaut est 200.</p> <p>Remarque: Pour un service Web, cette propriété affecte le nombre de demandes de service Web que le service d'intégration de données accepte avant d'envoyer les demandes vers le journal des travaux en souffrance du service d'intégration de données.</p>

Le tableau suivant décrit la propriété de taille maximum du tas que vous pouvez configurer pour le processus du service d'intégration de données :

Propriété	Description
Taille maximale du tas mémoire	Quantité de RAM allouée à la machine virtuelle Java (JVM) qui exécute le service d'intégration de données. Utilisez cette propriété pour optimiser les performances. Ajoutez l'une des lettres suivantes à la valeur afin de spécifier les unités : <ul style="list-style-type: none"> - o pour octets. - k pour kilo-octets. - m pour mégaoctets. - g pour gigaoctets. La valeur par défaut est 512 mégaoctets. Remarque: Pensez à augmenter la taille du tas mémoire lorsque le service d'intégration de données doit traiter de grandes quantités de données.

Exemple de configuration du service d'intégration de données pour des demandes de service Web simultanées

Lorsque vous configurez la façon dont le service d'intégration de données traite les demandes de services Web simultanées, vérifiez que la valeur du nombre maximal de demandes simultanées est la même pour le service Web et pour le processus du service d'intégration de données.

Par exemple, dans la configuration suivante, le service d'intégration de données accepte 200 demandes HTTP simultanées, mais seulement 10 demandes de service Web simultanées :

Type de propriété	Nom de la propriété	Configuration
Processus du service d'intégration de données	Nombre maximal de demandes simultanées	200
Processus du service d'intégration de données	Nombre maximal de demandes en retard	500
Service d'intégration de données	Taille maximale du pool d'exécution	100
Service Web	Nombre maximal de demandes simultanées	10

Quand le service d'intégration de données reçoit 20 demandes de service Web, 10 demandes de service Web échouent, car le service Web ne peut accepter que 10 demandes simultanées.

Pour éviter que les demandes de service Web échouent quand le service Web atteint son nombre maximal de demandes simultanées, configurez la même valeur maximale pour le processus du service d'intégration de données et le service Web. Lorsque le nombre de demandes envoyées au service d'intégration de données dépasse la valeur du nombre maximal de demandes simultanées, les demandes supplémentaires restent en attente jusqu'à ce que le processus du service d'intégration de données soit disponible pour les traiter.

Propriété du service Web pour configurer une instance DTM active

Pour améliorer les performances, vous pouvez configurer le service d'intégration de données de manière à ce qu'il conserve une instance DTM active dans le but que cette dernière traite plusieurs demandes de service

Web. Vous pouvez configurer la propriété « Durée d'activation du DTM » du service d'intégration de données dans l'outil Administrator.

Le tableau suivant décrit la propriété « Durée d'activation du DTM » :

Propriété	Description
Durée d'activation du DTM	<p>Nombre de millisecondes pendant lesquelles l'instance DTM demeure ouverte après le traitement de la dernière demande. Les demandes de service Web émises pour la même opération peuvent réutiliser l'instance ouverte. Utilisez le délai keepalive pour améliorer les performances lorsque le délai requis pour traiter la demande est limité par rapport au délai d'initialisation de l'instance DTM. Si la demande échoue, l'instance DTM prend fin.</p> <p>La valeur par défaut est 5 000.</p> <p>Remarque: grâce à la possibilité d'utiliser une instance DTM existante, les performances sont améliorées. Le service d'intégration de données (DIS) requiert des ressources supplémentaires pour démarrer une instance DTM à chaque demande. Le fait de garder le DTM actif consomme de la mémoire. Par conséquent, les utilisateurs doivent tenir compte de la consommation de mémoire lorsqu'ils configurent cette option.</p>

Mise en cache de l'ensemble de résultats du service Web

Lorsque vous configurez la mise en cache de l'ensemble de résultats, le Data Integration Service met en cache les résultats du processus DTM associé à chaque requête du service Web. Le Data Integration Service met en cache les résultats de la période d'expiration que vous définissez. Quand un client externe effectue la même requête avant l'expiration du cache, le Data Integration Service renvoie les résultats mis en cache.

Envisagez la solution suivante pour les goulots d'étranglement de cache de l'ensemble de résultats :

Configurez le cache de l'ensemble de résultats pour un service Web.

La mise en cache de l'ensemble des résultats permet au Data Integration Service d'utiliser les résultats mis en cache pour les demandes du service Web. Les utilisateurs qui exécutent des requêtes identiques dans un laps de temps court souhaitent peut-être utiliser la mise en cache de l'ensemble de résultats pour diminuer le temps d'exécution des requêtes identiques.

Le Data Integration Service stocke le cache de l'ensemble de résultats des services Web par utilisateur lorsque le service Web utilise WSSecurity. Le Data Integration Service stocke le cache en fonction du nom d'utilisateur qui est fourni dans le nom d'utilisateur de la requête de service Web. Lorsque le Data Integration Service met en cache les résultats par utilisateur, le Data Integration Service renvoie seulement les résultats mis en cache à l'utilisateur qui a envoyé la requête du service Web.

Activation de la mise en cache de l'ensemble des résultats pour un service Web

Pour utiliser des résultats mis en cache pour des requêtes de service Web identiques, configurez le Data Integration Service pour qu'il utilise la mise en cache de l'ensemble des résultats.

1. Dans l'outil Administrator, sélectionnez un Data Integration Service.
2. Cliquez sur la vue **Processus** pour configurer les propriétés de mise en cache de l'ensemble des résultats.

3. Cliquez sur la vue **Application**, cliquez sur le service Web, puis cliquez sur l'opération pour configurer la période d'expiration du cache dans les propriétés d'opération du service Web. Pour que le Data Integration Service mette en cache les résultats par utilisateur, activez WS-Security dans les propriétés du service Web.
4. Pour désactiver la mise en cache de l'ensemble des résultats pour une requête de service Web lorsque l'opération du service Web est configurée pour mettre en cache l'ensemble des résultats, incluez la syntaxe suivante dans l'en-tête HTTP de la requête SOAP :

```
WebServiceOptions.disableResultSetCache=true
```

Gestion du journal du service Web

Les performances d'E/S du système peuvent diminuer lorsque le Data Integration Service écrit et conserve un grand nombre de fichiers journaux. Le Data Integration Service génère des journaux d'exécution du service Web basés sur le niveau de suivi que vous configurez. Pensez à gérer le nombre de fichiers journaux que le Data Integration Service écrit et conserve.

Envisagez les solutions suivantes pour les goulots d'étranglement de journal de service Web :

Définissez le niveau de suivi du service Web sur OFF.

Lorsque vous configurez les propriétés du service Web pour un service Web déployé, vous pouvez spécifier le niveau de suivi du journal. Le niveau de suivi détermine les types de journaux que le Data Integration Service écrit dans l'emplacement du journal d'exécution. Le niveau de suivi du service Web par défaut est INFO. Lorsque le niveau de suivi est défini sur FINEST ou ALL, la performance peut diminuer car le Data Integration Service écrit davantage de journaux dans le fichier journal. La définition du niveau de suivi sur FINEST ou ALL a un plus grand impact sur les performances lorsque le service Web utilise HTTPS et WS-Security.

Archivez les fichiers journaux dont vous n'avez plus besoin.

Les E/S du système sont affectées par le stockage d'un trop grand nombre de fichiers journaux. Par défaut, le Data Integration Service écrit les journaux d'exécution du service Web dans le répertoire suivant : `<InformaticaInstallationDir>/tomcat/bin/disLogs/ws`

Remarque: Si vous supprimez le dossier ws pour vider les journaux, vous devez le recréer. Arrêtez le Data Integration Service avant de supprimer et de recréer le dossier ws.

CHAPITRE 10

Optimisation des connexions

Ce chapitre comprend les rubriques suivantes :

- [Présentation de l'optimisation de connexions, 91](#)
- [Pooling de connexions, 91](#)
- [Taille du paquet réseau de base de données, 93](#)

Présentation de l'optimisation de connexions

Vous pouvez optimiser les connexions pour améliorer les performances. Vous pouvez gérer le pool d'instances de connexions inactives pour une connexion de base de données. Vous pouvez augmenter la taille du paquet réseau pour permettre à de plus grands paquets de données de traverser le réseau simultanément.

Utilisez les techniques suivantes pour optimiser les connexions :

- Optimisation du pooling de connexions.
- Optimisation de la taille de paquet réseau de base de données.

Pooling de connexions

Le pooling de connexions est une infrastructure permettant de mettre en cache les informations de connexion de la base de données qui sont utilisées par le service d'intégration de données. Il permet d'optimiser les performances grâce à la réutilisation d'informations de connexion mises en cache.

Envisagez la solution suivante en cas de goulots d'étranglement au niveau des connexions :

Activer le pooling de connexions pour une connexion de base de données.

Activer le pooling de connexions pour optimiser les performances de connexion. Vous pouvez gérer les instances de connexions inactives d'une connexion de base de données. Le pool de connexions conserve les instances de connexions inactives en se basant sur les propriétés de pooling que vous configurez. Vous pouvez indiquer le nombre maximal et minimal de connexions inactives et le temps d'attente maximal pour une connexion inactive.

Propriétés de pooling dans les objets de connexion

Vous pouvez modifier les propriétés du pooling de connexions dans la vue **Pooling** d'une connexion de base de données.

Le nombre de bibliothèques du pool de connexions dépend du nombre de processus de service d'intégration de données ou de processus DTM en cours d'exécution. Chaque processus du service d'intégration de données ou DTM conserve sa propre bibliothèque du pool de connexions. Les valeurs des propriétés du pooling concernent chaque bibliothèque du pool de connexions.

Par exemple, si vous définissez le nombre maximal de connexions sur 15, chaque bibliothèque du pool de connexions peut avoir au maximum 15 connexions inactives dans le pool. Si le service d'intégration de données exécute des tâches dans des processus locaux séparés et que trois processus DTM sont en cours d'exécution, vous pouvez avoir un maximum de 45 instances de connexions inactives.

Pour réduire le nombre total d'instances de connexions inactives, définissez le nombre minimal de connexions sur 0 et réduisez la durée d'inactivité maximale pour chaque connexion de base de données.

Le tableau suivant décrit les propriétés de pooling de connexions de base de données que vous pouvez modifier dans la vue **Pooling** pour une connexion de base de données :

Activer le pooling de connexions

Active le pooling de connexions. Lorsque vous activez le pooling de connexions, chaque pool de connexions conserve les instances de connexion inactives en mémoire. Pour supprimer le pool de connexions inactives, vous devez redémarrer le service d'intégration de données.

Si le pooling de connexion est désactivé, le processus DTM ou le processus du service d'intégration de données arrête toutes les activités de pooling. Le processus DTM ou le processus du service d'intégration de données crée une instance de connexion à chaque fois qu'il traite une tâche. Il annule l'instance lorsqu'il termine le traitement de la tâche.

Par défaut, activée pour les connexions DB2 for i5/OS, DB2 for z/OS, IBM DB2, Microsoft SQL Server, Oracle et ODBC. Par défaut, désactivée pour les connexions Adabas, IMS, Sequential et VSAM.

Le paramètre par défaut est activé pour les connexions Microsoft SQL Server, IBM DB2, Oracle et ODBC.

Nombre minimum de connexions

Nombre minimal d'instances de connexions inactives conservé par un pool pour une connexion de base de données une fois la durée maximale d'inactivité atteinte. Définissez cette valeur qui doit être inférieure ou égale au nombre maximal d'instances de connexions inactives. La valeur par défaut est 0.

Nombre maximum de connexions

Nombre maximal d'instances de connexions inactives conservé par un pool pour une connexion de base de données avant que la durée maximale d'inactivité ne soit atteinte. Définissez cette valeur à une valeur supérieure au nombre minimum d'instances de connexion inactives. La valeur par défaut est 15.

Durée maximum d'inactivité

Nombre de secondes durant lesquelles une instance de connexion qui dépasse le nombre minimal d'instances de connexion peut rester inactive avant que le pool de connexions ne l'annule. Le pool de connexions ignore la durée d'inactivité lorsque l'instance de connexion ne dépasse pas le nombre minimal d'instances de connexions inactives. La valeur par défaut est 120.

Taille du paquet réseau de base de données

Si vous lisez à partir de cibles Oracle, Sybase ASE ou Microsoft SQL Server ou écrivez dedans, vous pouvez améliorer les performances en augmentant la taille du paquet réseau en fonction de la base de données à partir de laquelle vous lisez ou dans laquelle vous écrivez. En augmentant la taille du paquet réseau, vous permettez à de plus grands paquets de données de traverser le réseau simultanément.

Envisagez les solutions suivantes pour les goulots d'étranglement de taille du paquet réseau de base de données :

Augmentez la taille du paquet réseau de base de données pour une base de données Oracle.

Vous pouvez augmenter la taille du paquet réseau de serveur de base de données dans listener.ora et tnsnames.ora. Consultez la documentation de votre base de données pour de plus amples informations sur l'augmentation de la taille de paquet, si besoin.

Augmentez la taille du paquet réseau de base de données pour une base de données Sybase ASE.

Consultez la documentation de votre base de données pour savoir comment augmenter la taille du paquet. Vous devez également changer la taille du paquet pour Sybase ASE dans l'objet de connexion relationnel dans le Data Integration Service pour refléter la taille du paquet de serveur de base de données.

Augmentez la taille du paquet réseau de base de données pour une base de données Microsoft SQL Server.

Consultez la documentation de votre base de données pour savoir comment augmenter la taille du paquet. Vous devez également changer la taille du paquet pour Microsoft SQL Server dans l'objet de connexion relationnel dans le Data Integration Service pour refléter la taille du paquet de serveur de base de données.

INDEX

A

- A des effets secondaires
 - description de propriété de transformation [39](#)
- activation de la mise en cache de l'ensemble des résultats pour un service de données SQL
 - cache de l'ensemble des résultats [81](#)
- allocation de mémoire
 - demandes simultanées [79](#)
 - instance DTM active [89](#)
 - Service de données SQL [79](#)
 - Service Web [89](#)

B

- bases de données
 - optimisation des cibles pour le partitionnement [61](#)
 - optimisation des sources pour le partitionnement [61](#)

C

- cache d'objet de données
 - cache d'index [70](#)
 - configuration [70](#)
 - description [70](#)
 - optimisation [72](#)
 - tables gérées par l'utilisateur [70](#)
- cache de l'ensemble de résultats de service de données SQL
 - Service d'intégration de données [80](#)
- cache de l'ensemble de résultats du service Web
 - Data Integration Service [89](#)
- cache de l'ensemble des résultats
 - activation de la mise en cache de l'ensemble des résultats pour un service de données SQL [81](#)
 - propriétés du cache de l'ensemble de résultats [81](#)
- cache de transformation
 - optimisation de transformation [38](#)
- chargements en bloc
 - optimisation de cible [17](#)
- cible de fichier plat
 - optimisation de cible [16](#)
- compression de messages du service Web
 - optimisation du service Web [84](#)
- contraintes
 - configuration de contraintes [24](#)
 - optimisation de source [23](#)

D

- Data Integration Service
 - cache de l'ensemble de résultats du service Web [89](#)
- demandes simultanées
 - Service de données SQL [79](#)

- demandes simultanées (*a continué*)
 - Service Web [86](#)

E

- effets secondaires
 - description [39](#)
 - Transformation Consommateur de service Web [41](#)
 - transformation SQL [37](#)
- élimination d'erreurs de transformation
 - optimisation de transformation [39](#)

F

- fichiers plats
 - optimisation des cibles pour le partitionnement [60](#)
 - optimisation des sources pour le partitionnement [60](#)
- filtres conditionnels
 - optimisation de source [21](#)

G

- gestion du journal du service Web
 - niveau de suivi d'erreur [90](#)
- goulots d'étranglement
 - sous Unix [14](#)
 - sous Windows [13](#)

I

- indicateurs
 - Vue Requête [23](#)
- indicateurs de base de données
 - saisie dans l'outil Developer [23](#)
- indicateurs SQL
 - saisie dans l'outil Developer [23](#)
- instance DTM active
 - Service Web [89](#)
- intervalles de point de contrôle de base de données
 - optimisation de cible [17](#)

L

- lecture à un seul passage
 - optimisation du mappage [54](#)

M

- mappages
 - méthode d'optimisation de prédicat [47](#)

- mappages (*a continué*)
 - méthode d'optimisation globale des prédicats [51](#)
 - méthodes d'optimisation [45](#)
 - optimisation du partitionnement [58](#)
- méthode d'optimisation du refoulement
 - refoulement complet [53](#)
 - refoulement de la source [54](#)
- mise en cache d'objet de données
 - types de données de table [70](#)

N

- niveau d'optimisation complet
 - description [45](#)
- niveau d'optimisation minimal
 - description [45](#)
- niveau d'optimisation normal
 - description [45](#)
- niveau de suivi d'erreur
 - gestion du journal du service Web [90](#)
- niveaux d'optimisation
 - description [45](#)

O

- objet de données personnalisé
 - optimisation de source [24](#)
- objets de données logiques
 - mise en cache dans la base de données [70](#)
- optimisation
 - effets secondaires [39](#)
 - méthode d'optimisation basée sur le coût [48](#)
 - méthode d'optimisation de jointure dataship [49](#)
 - méthode d'optimisation de nettoyage de branche [51](#)
 - Méthode d'optimisation de projection précoce [46](#)
 - méthode d'optimisation de sélection précoce [51](#)
 - méthode d'optimisation de semi-jointure [50](#)
 - méthode d'optimisation du refoulement [52](#)
 - méthode d'optimisation push-into [52](#)
 - méthodes de performance de mappage [45](#)
- optimisation basée sur le coût
 - description [48](#)
- optimisation d'expression
 - optimisation du mappage [28](#)
- optimisation de base de données Oracle
 - optimisation de cible [18](#)
 - optimisation de source [25](#)
- optimisation de cible
 - chargements en bloc [17](#)
 - cible de fichier plat [16](#)
 - intervalles de point de contrôle de base de données [17](#)
 - optimisation de base de données Oracle [18](#)
- optimisation de conversion du type de données
 - optimisation du mappage [56](#)
- optimisation de filtre
 - optimisation du mappage [55](#)
- optimisation de jointure dataship
 - description [49](#)
- optimisation de nettoyage de branche
 - description [51](#)
- optimisation de projection précoce
 - description [46](#)
- optimisation de requête
 - optimisation de source [20](#)
- optimisation de sélection précoce
 - description [51](#)

- optimisation de sélection précoce (*a continué*)
 - Transformation Consommateur de service Web [41](#)
 - transformation SQL [37](#)
- optimisation de semi-jointure
 - description [50](#)
- optimisation de source
 - contraintes [23](#)
 - filtres conditionnels [21](#)
 - objet de données personnalisé [24](#)
 - optimisation de base de données Oracle [25](#)
 - optimisation de requête [20](#)
 - Sélectionner distinct [21](#)
 - source de fichier plat [20](#)
- optimisation de transformation
 - cache de transformation [38](#)
 - élimination d'erreurs de transformation [39](#)
 - Transformation Agrégation [27](#)
 - Transformation Consommateur de service Web [40](#)
 - transformation Java [30](#)
 - Transformation Jointure [32](#)
 - transformation Processeur de données [27](#)
 - Transformation Recherche [33](#)
 - Transformation SQL [36](#)
 - transformation Trieur [36](#)
- optimisation des connexions
 - pooling de connexions [91](#)
 - taille du paquet réseau de base de données [93](#)
- optimisation des requêtes HTTP
 - optimisation du service Web [84](#)
- optimisation du mappage
 - lecture à un seul passage [54](#)
 - optimisation d'expression [28](#)
 - optimisation de conversion du type de données [56](#)
 - optimisation de filtre [55](#)
 - suivi d'erreur [56](#)
- optimisation du Model Repository Service
 - optimisation du temps d'exécution [66](#)
- optimisation du refoulement
 - description [52](#)
- optimisation du service Analyst
 - optimisation du temps d'exécution [65](#)
- Optimisation du service d'intégration de données
 - optimisation du temps d'exécution [65](#)
- optimisation du service de données SQL
 - outils client tiers [75](#)
 - pilotes JDBC [75](#)
- optimisation du service Web
 - compression de messages du service Web [84](#)
 - optimisation des requêtes HTTP [84](#)
- optimisation du système
 - optimisation du temps d'exécution [73](#)
- optimisation du temps d'exécution
 - optimisation du Model Repository Service [66](#)
 - optimisation du service Analyst [65](#)
 - Optimisation du service d'intégration de données [65](#)
 - optimisation du système [73](#)
 - statistiques de surveillance [67](#)
- optimisation push-into
 - Activation dans la transformation SQL [38](#)
 - description [52](#)
 - Transformation Consommateur de service Web [41](#)
 - Transformation SQL [37](#)
- outils client tiers
 - optimisation du temps d'exécution [75](#)

P

- parallélisme maximal
 - augmentation [59](#)
- partitionnement
 - optimisation [58](#)
 - optimisation des bases de données cible [61](#)
 - optimisation des bases de données source [61](#)
 - optimisation des cibles de fichiers plats [60](#)
 - optimisation des sources de fichiers plats [60](#)
 - optimisation des transformations [62](#)
 - plusieurs processeurs [59](#)
- pilotes JDBC
 - optimisation du temps d'exécution [75](#)
- Plans de requête SQL
 - affichage [78](#)
- pooling de connexions
 - optimisation des connexions [91](#)
 - propriétés [92](#)
- port de filtre
 - Transformation Consommateur de service Web [41](#)
- propriétés du cache de l'ensemble de résultats
 - optimisation du temps d'exécution [81](#)

R

- réglage de performance
 - méthode d'optimisation de jointure dataship [49](#)
 - méthode d'optimisation de nettoyage de branche [51](#)
 - méthode d'optimisation de semi-jointure [50](#)
 - méthode d'optimisation du refoulement [52](#)
 - méthode d'optimisation globale des prédicats [51](#)
 - méthodes d'optimisation [45](#)
- réglage des performances
 - méthode d'optimisation basée sur le coût [48](#)
 - méthode d'optimisation de prédicat [47](#)
 - Méthode d'optimisation de projection précoce [46](#)
 - méthode d'optimisation de sélection précoce [51](#)
 - méthode d'optimisation push-into [52](#)
 - niveaux d'optimisation [45](#)

S

- Sélectionner distinct
 - optimisation de source [21](#)
- Service d'intégration de données
 - cache de l'ensemble de résultats de service de données SQL [80](#)
- Service de données SQL
 - allocation de mémoire [79](#)
- Service Web
 - allocation de mémoire [89](#)
 - demandes simultanées [86](#)
- source de fichier plat
 - optimisation de source [20](#)
- statistiques de surveillance
 - optimisation du temps d'exécution [67](#)
- suivi d'erreur
 - optimisation du mappage [56](#)
- système
 - goulots d'étranglement sous UNIX, identification [14](#)

- système (*a continué*)
 - goulots d'étranglement sous Windows, identification [13](#)

T

- tables temporaires
 - description [82](#)
- tables virtuelles
 - mise en cache dans la base de données [70](#)
- taille du paquet réseau de base de données
 - optimisation des connexions [93](#)
- Transformation Agrégation
 - optimisation de transformation [27](#)
- Transformation Consommateur de service Web
 - Activation de l'optimisation push-into [42](#)
 - optimisation de filtre [41](#)
 - optimisation de sélection précoce [41](#)
 - optimisation de transformation [40](#)
 - optimisation push-into [41](#)
- transformation Java
 - optimisation de transformation [30](#)
- Transformation Jointure
 - optimisation de transformation [32](#)
- transformation Processeur de données
 - optimisation de transformation [27](#)
- Transformation Recherche
 - optimisation de transformation [33](#)
- transformation SQL
 - optimisation de sélection précoce [37](#)
- Transformation SQL
 - optimisation de transformation [36](#)
 - optimisation push-into [37](#)
 - propriétés de l'optimisation push-into [38](#)
- transformation Trieur
 - optimisation de transformation [36](#)
- transformations
 - optimisation pour le partitionnement [62](#)

U

- UNIX
 - goulots d'étranglement du système [14](#)

V

- Vue Requête
 - configuration d'indicateurs [23](#)

W

- Windows
 - goulots d'étranglement [13](#)