Informatica

Informatica®
10.1.0

Leistungsoptimierungs-Handbuch

Informatica Leistungsoptimierungs-Handbuch
10.1.0
Juni 2016

© Copyright Informatica LLC 1998, 2018

Diese Software und die zugehörige Dokumentation enthalten proprietäre Informationen der Informatica LLC, werden unter einem Lizenzvertrag mit Einschränkungen hinsichtlich Verwendung und Veröffentlichung zur Verfügung gestellt und sind urheberrechtlich geschützt. Das Zurückentwickeln (Reverse Engineering) der Software ist untersagt. Ohne ausdrückliche schriftliche Genehmigung der Informatica LLC darf kein Teil dieses Dokuments zu irgendeinem Zweck vervielfältigt oder übertragen werden, unabhängig davon, auf welche Art und Weise oder mit welchen Mitteln (elektronisch, mechanisch, durch Fotokopieren, Aufzeichnen usw.) dies geschieht. Diese Software ist möglicherweise durch US-amerikanische und/oder internationale Patente und weitere angemeldete Patente geschützt.

Die Verwendung, Vervielfältigung oder Veröffentlichung der Software durch die US-Regierung unterliegt den Bestimmungen des jeweiligen Softwarelizenzvertrags sowie ggf. den Bestimmungen in DFARS 227.7202-1(a) und 227.7702-3(a) (1995), DFARS 252.227-7013 © (1)(ii) (OCT. 1988), FAR 12.212(a) (1995), FAR 52.227-19 oder FAR 52.227-14 (ALT III).

Die in diesem Produkt und in dieser Dokumentation enthaltenen Informationen können jederzeit ohne vorherige Ankündigung geändert werden. Sollten Sie mit diesem Produkt oder dieser Dokumentation Probleme haben, teilen Sie uns dies bitte schriftlich mit.

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, Informatica Master Data Management und Live Data Map sind Marken oder eingetragene Marken der Informatica LLC in den USA und anderen Ländern. Alle weiteren Produkt- und Firmennamen sind möglicherweise Markennamen oder Warenzeichen der jeweiligen Eigentümer.

Teile dieser Software und/oder Dokumentation sind durch die Urheberrechte Dritter geschützt, einschließlich und ohne Einschränkung: Copyright DataDirect Technologies. Alle Rechte vorbehalten. Copyright © Sun Microsystems. Alle Rechte vorbehalten. Copyright © RSA Security Inc. Alle Rechte vorbehalten. Copyright © Ordinal Technology Corp. Alle Rechte vorbehalten. Copyright © Aandacht c.v. Alle Rechte vorbehalten. Copyright Genivia, Inc. Alle Rechte vorbehalten. Copyright Isomorphic Software. Alle Rechte vorbehalten. Copyright © Meta Integration Technology, Inc. Alle Rechte vorbehalten. Copyright © Intalio. Alle Rechte vorbehalten. Copyright © Oracle. Alle Rechte vorbehalten. Copyright © Adobe Systems Incorporated. Alle Rechte vorbehalten. Copyright © DataArt, Inc. Alle Rechte vorbehalten. Copyright © ComponentSource. Alle Rechte vorbehalten. Copyright © Microsoft Corporation. Alle Rechte vorbehalten. Copyright © Rouge Wave Software, Inc. Alle Rechte vorbehalten. Copyright © Teradata Corporation. Alle Rechte vorbehalten. Copyright © Yahoo! Inc. Alle Rechte vorbehalten. Copyright © Glyph & Cog, LLC. Alle Rechte vorbehalten. Copyright © Thinkmap, Inc. Alle Rechte vorbehalten. Copyright © Clearpace Software Limited. Alle Rechte vorbehalten. Copyright © Information Builders, Inc. Alle Rechte vorbehalten. Copyright © OSS Nokalva, Inc. Alle Rechte vorbehalten. Copyright Edifecs, Inc. Alle Rechte vorbehalten. Copyright Cleo Communications, Inc. Alle Rechte vorbehalten. Copyright © International Organization for Standardization 1986. Alle Rechte vorbehalten. Copyright © ej-technologies GmbH . Alle Rechte vorbehalten. Copyright © Jaspersoft Corporation. Alle Rechte vorbehalten. Copyright © International Business Machines Corporation. Alle Rechte vorbehalten. Copyright © yWorks GmbH. Alle Rechte vorbehalten. Copyright © Lucent Technologies. Alle Rechte vorbehalten. Copyright © Universität von Toronto. Alle Rechte vorbehalten. Copyright © Daniel Veillard. Alle Rechte vorbehalten. Copyright © Unicode, Inc. Copyright IBM Corp. Alle Rechte vorbehalten. Copyright © MicroQuill Software Publishing, Inc. Alle Rechte vorbehalten. Copyright © PassMark Software Pty Ltd. Alle Rechte vorbehalten. Copyright © LogiXML, Inc. Alle Rechte vorbehalten. Copyright © 2003-2010 Lorenzi Davide. Alle Rechte vorbehalten. Copyright © Red Hat, Inc. Alle Rechte vorbehalten. Copyright © The Board of Trustees of the Leland Stanford Junior University. Alle Rechte vorbehalten. Copyright © EMC Corporation. Alle Rechte vorbehalten. Copyright © Flexera Software. Alle Rechte vorbehalten. Copyright © Jinfonet Software. Alle Rechte vorbehalten. Copyright © Apple Inc. Alle Rechte vorbehalten. Copyright © Telerik Inc. Alle Rechte vorbehalten. Copyright © BEA Systems. Alle Rechte vorbehalten. Copyright © PDFlib GmbH. Alle Rechte vorbehalten. Copyright © Orientation in Objects GmbH. Alle Rechte vorbehalten. Copyright © Tanuki Software, Ltd. Alle Rechte vorbehalten. Copyright © Ricebridge. Alle Rechte vorbehalten. Copyright © Sencha, Inc. Alle Rechte vorbehalten. Copyright © Scalable Systems, Inc. Alle Rechte vorbehalten. Copyright © jQWidgets. Alle Rechte vorbehalten. Copyright © Tableau Software, Inc. Alle Rechte vorbehalten. Copyright© MaxMind, Inc. Alle Rechte vorbehalten. Copyright © TMate Software s.r.o. Alle Rechte vorbehalten. Copyright © MapR Technologies Inc. Alle Rechte vorbehalten. Copyright © Amazon Corporate LLC. Alle Rechte vorbehalten. Copyright © Highsoft. Alle Rechte vorbehalten. Copyright © Python Software Foundation. Alle Rechte vorbehalten. Copyright © BeOpen.com. Alle Rechte vorbehalten. Copyright © CNRI. Alle Rechte vorbehalten.

Dieses Produkt enthält Software, die von der Apache Software Foundation (http://www.apache.org/) entwickelt wurde, und andere Software, die unter den Bedingungen des Apache-Lizenzvertrags lizenziert ist („Lizenz“). Eine Kopie dieser Lizenzen finden Sie unter http://www.apache.org/licenses/. Sofern nicht gesetzlich vorgeschrieben oder schriftlich vereinbart, erfolgt der Vertrieb der Software unter der Lizenz auf der BASIS „WIE BESEHEN” OHNE GARANTIEN ODER KONDITIONEN IRGENDEINER ART, weder ausdrücklich noch impliziert. Berechtigungen und Einschränkungen für bestimmte Sprachen finden Sie in der Lizenz.

Dieses Produkt enthält Software, die von Mozilla (http://www.mozilla.org/) entwickelt wurde, Software Copyright The JBoss Group, LLC. Alle Rechte vorbehalten; Software Copyright © 1999-2006 by Bruno Lowagie und Paulo Soares, und andere Software, die gemäß den verschiedenen Versionen des GNU Lesser General Public License Agreement unter http://www.gnu.org/licenses/lgpl.html lizenziert ist. Die Materialien werden „wie besehen“ kostenlos von Informatica bereitgestellt, ohne ausdrückliche oder stillschweigende Gewährleistung, einschließlich, jedoch nicht beschränkt auf die stillschweigenden Gewährleistungen der Handelsüblichkeit und der Eignung für einen bestimmten Zweck.

Das Produkt enthält ACE(TM) und TAO(TM) Software, Copyright Douglas C. Schmidt und seine Forschungsgruppe an der Washington University, University of California, Irvine und Vanderbilt University, Copyright (©) 1993-2006. Alle Rechte vorbehalten.

Dieses Produkt enthält Software, die von OpenSSL Project zur Verwendung im OpenSSL Toolkit entwickelt wurde (Copyright The OpenSSL Project. Alle Rechte vorbehalten). Die erneute Verteilung dieser Software unterliegt den unter „http://www.openssl.org“ und „http://www.openssl.org/source/license.html“ verfügbaren Bedingungen.

Dieses Produkt enthält urheberrechtlich geschützte Curl-Software (Copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>). Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://curl.haxx.se/docs/copyright.html“ verfügbaren Bedingungen. Die Erlaubnis, diese Software für jeden beliebigen Zweck gegen Gebühr oder kostenlos zu verwenden, zu kopieren, zu ändern und zu verteilen, wird hiermit erteilt, sofern die oben genannten urheberrechtlichen Hinweise und diese Erlaubnis in allen Exemplaren angegeben werden.

Das Produkt enthält urheberrechtlich geschützte Software, Copyright 2001-2005 (©) MetaStuff, Ltd. Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://www.dom4j.org/ license.html“ verfügbaren Bedingungen.

Das Produkt enthält urheberrechtlich geschützte Software, Copyright © 2004-2007, The Dojo Foundation. Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://dojotoolkit.org/license“ verfügbaren Bedingungen.

Dieses Produkt enthält urheberrechtlich geschützte ICU-Software, Copyright International Business Machines Corporation und andere. Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://source.icu-project.org/repos/icu/icu/trunk/license.html“ verfügbaren Bedingungen.

Dieses Produkt enthält urheberrechtlich geschützte Software, Copyright © 1996-2006 Per Bothner. Alle Rechte vorbehalten. Das Ihnen erteilte Recht, diese Materialien zu verwenden, unterliegt den unter „http:// www.gnu.org/software/ kawa/Software-License.html“ verfügbaren Bedingungen.

Dieses Produkt enthält urheberrechtlich geschützte OSSP UUID-Software (Copyright © 2002 Ralf S. Engelschall, Copyright © 2002 The OSSP Project Copyright © 2002 Cable & Wireless Deutschland). Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://www.opensource.org/licenses/mit-license.php“ verfügbaren Bedingungen.

Dieses Produkt enthält Software, die von Boost (http://www.boost.org/) oder unter der Softwarelizenz von Boost entwickelt wurde. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://www.boost.org/LICENSE_1_0.txt“ verfügbaren Bedingungen.

Dieses Produkt enthält urheberrechtlich geschützte Software, Copyright © 1997-2007 University of Cambridge. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter http://www.pcre.org/license.txt einsehbaren Bedingungen.

Dieses Produkt enthält urheberrechtlich geschützte Software, Copyright © 2007 The Eclipse Foundation. Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://www.eclipse.org/org/documents/epl-v10.php“ und „http://www.eclipse.org/org/documents/edl-v10.php“ verfügbaren Bedingungen.

Dieses Produkt enthält Software gemäß den Lizenzbedingungen unter 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/hsqlLicense.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; https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt.

Dieses Produkt enthält Software, die unter der Academic Free License (http://www.opensource.org/licenses/afl-3.0.php), der Common Development Distribution License (http://www.opensource.org/licenses/cddl1.php), der Common Public License (http://www.opensource.org/licenses/cpl1.0.php), den Sun Binary Code License Agreement Supplemental License Terms, der BSD License (http:// www.opensource.org/licenses/bsd-license.php), der neuen BSD License (http://opensource.org/licenses/BSD-3-Clause), der MIT License (http://www.opensource.org/licenses/mit-license.php), der Artistic License (http://www.opensource.org/licenses/artistic-license-1.0) und der Initial Developer's Public License Version 1.0 (http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/) lizenziert ist.

Dieses Produkt enthält urheberrechtlich geschützte Software, Copyright © 2003-2006 Joe Walnes, 2006-2007 XStream Committers. Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://xstream.codehaus.org/license.html“ verfügbaren Bedingungen. Dieses Produkt enthält Software, die von der Indiana University Extreme! Lab. entwickelt wurde. Weitere Informationen finden Sie unter http://www.extreme.indiana.edu/.

Dieses Produkt enthält Software, Copyright © 2013 Frank Balluffi und Markus Moeller. Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den Bedingungen der MIT-Lizenz.

Weitere Informationen über die Patente finden Sie unter https://www.informatica.com/legal/patents.html.

HAFTUNGSAUSSCHLUSS: Informatica LLC stellt diese Dokumentation „wie besehen“ bereit, ohne ausdrückliche oder stillschweigende Gewährleistung, einschließlich, jedoch nicht beschränkt auf die Gewährleistungen der Nichtverletzung der Rechte von Dritten, der Handelsüblichkeit oder Eignung für einen bestimmten Zweck. Informatica LLC garantiert nicht die Fehlerfreiheit dieser Software oder Dokumentation. Die in dieser Software oder Dokumentation bereitgestellten Informationen können technische Ungenauigkeiten oder Druckfehler enthalten. Die in dieser Software und in dieser Dokumentation enthaltenen Informationen können jederzeit ohne vorherige Ankündigung geändert werden.

HINWEISE

Dieses Informatica-Produkt (die „Software“) umfasst bestimmte Treiber (die „DataDirect-Treiber“) von DataDirect Technologies, einem Betreiber von Progress Software Corporation („DataDirect“), die folgenden Bedingungen und Bestimmungen unterliegen:

1. DIE DATADIRECT-TREIBER WERDEN „WIE GESEHEN“ OHNE JEGLICHE GEWÄHRLEISTUNG, WEDER AUSDRÜCKLICH NOCH STILLSCHWEIGEND, BEREITGESTELLT, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE STILLSCHWEIGENDEN GEWÄHRLEISTUNGEN DER HANDELSÜBLICHKEIT, EIGNUNG FÜR EINEN BESTIMMTEN ZWECK UND DER NICHTVERLETZUNG VON RECHTEN DRITTER.
2. IN KEINEM FALL SIND DATADIRECT ODER DRITTANBIETER DEM ENDBENUTZER GEGENÜBER HAFTBAR FÜR UNMITTELBARE, MITTELBARE, KONKRETE, NEBEN-, FOLGE- ODER ANDERE SCHÄDEN, DIE SICH AUS DER VERWENDUNG DER ODBC-TREIBER ERGEBEN, UNABHÄNGIG DAVON, OB SIE IM VORAUS ÜBER DIE MÖGLICHKEIT SOLCHER SCHÄDEN INFORMIERT WORDEN SIND ODER NICHT. DIESE BESCHRÄNKUNGEN GELTEN FÜR ALLE KLAGEGEGENSTÄNDE, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF VERTRAGSBRUCH, GEWÄHRLEISTUNGSBRUCH, FAHRLÄSSIGKEIT, KAUSALHAFTUNG, TÄUSCHUNG UND ANDERE UNERLAUBTE HANDLUNGEN.

Publikationsdatum: 2018-07-23

Inhalt

Einleitung 8

Informatica-Ressourcen. 8

Informatica-Netzwerk. 8

Informatica-Wissensdatenbank. 8

Informatica-Dokumentation. 9

Informatica-Produktverfügbarkeitsmatrizen. 9

Informatica Velocity. 9

Informatica Marketplace. 9

Globaler Kundensupport von Informatica. 9

Kapitel 1: Leistungsoptimierung - Übersicht. 10

Leistungsoptimierung - Übersicht. 10

Target-Engpässe. 11

Quellengpässe. 11

Mapping-Engpässe. 12

Engpässe im Computersystem. 12

Bestimmen von Systemengpässen unter Windows. 12

Bestimmen von Systemengpässen unter UNIX. 13

Laufzeit-Engpässe. 13

Engpässe bei der SQL-Datendienstoptimierung. 13

Engpässe bei der Web-Dienst-Optimierung. 14

Verbindungsengpässe. 14

Kapitel 2: Target-Optimierung. 15

Ziel-Optimierung - Übersicht. 15

Optimierung von Einfachdateizielen. 15

Datenbankprüfpunkt-Intervalle. 16

Massenladevorgänge. 16

Datenbank-Ziel-Optimierung. 17

Kapitel 3: Quelloptimierung. 18

Quelloptimierung - Übersicht. 18

Optimierung von Einfachdateiquellen. 19

Abfrageoptimierung. 19

Bedingungsfilter. 20

Select Distinct. 20

Hinweise. 21

Regeln und Richtlinien für Hinweise. 21

Erstellen von Hinweisen. 22

Einschränkungen. 22

Konfigurieren von Einschränkungen. . . . 23
Optimierung von konfigurierten Datenobjekten. . . . 23
Datenbankquelloptimierung. . . . 24

Kapitel 4: Umwandlungsoptimierung. . . . 25

Umwandlungsoptimierung. . . . 25
Aggregat-Umwandlungsoptimierung. . . . 25
Expressionsoptimierung. . . . 26
Java-Umwandlungsoptimierung. . . . 28
Frühe Auswahloptimierung mit der Java-Umwandlung. . . . 28
Push-Into-Optimierung mit der Java-Umwandlung. . . . 30
Joiner-Umwandlungsoptimierung. . . . 31
Lookup-Umwandlungsoptimierung. . . . 32
Sortierer-Umwandlungsoptimierung. . . . 34
SQL-Umwandlungsoptimierung. . . . 35
"Early Selection"-Optimierung mit der SQL-Umwandlung. . . . 35
Push-Into-Optimierung mit der SQL-Umwandlung. . . . 36
Umwandlungs-Cache. . . . 37
Eliminierung von Umwandlungsfehlern. . . . 37
Nebenwirkungen von Umwandlungen. . . . 38
Optimierung der Webdienstverbraucher-Umwandlung. . . . 39
Frühe Auswahloptimierung mit der Webdienstkunden-Umwandlung. . . . 39
Push-Into-Optimierung mit der Web-Dienst-Verbraucher-Umwandlung. . . . 40

Kapitel 5: Mapping-Optimierung. . . . 42

Mapping-Optimierung - Übersicht. . . . 42
Optimierungsmethoden. . . . 43
Optimierungsebenen. . . . 43
Filteroptimierungen. . . . 44
Frühe Projektionsoptimierungsmethode. . . . 44
Vorhersageoptimierungsmethode. . . . 45
Kostenbasierte Optimierungsmethode. . . . 46
Dataship-Join-Optimierungsmethode. . . . 46
Semi-Join Optimierungsmethode. . . . 47
Frühe Auswahloptimierungsmethode. . . . 48
Globale Vorhersageoptimierungsmethode. . . . 49
Methode zur Optimierung der Verzweigungsbereinigung. . . . 49
Push-Into-OptimierungMethode. . . . 49
Pushdown-Optimierung. . . . 50
Vollständige Pushdown-Optimierung. . . . 50
Quell-Pushdown. . . . 51
Regeln und Richtlinien für die Pushdown-Optimierung. . . . 52
Lesen in einem einzigen Durchlauf. . . . 52

Filteroptimierung. 53
Optimierung der Datentypkonvertierung. 53
Fehler-Tracing. 54

Kapitel 6: Optimierung durch partitionierte Mappings. 55

Optimierung durch partitionierte Mappings - Übersicht. 55
Verwenden von mehreren CPUs. 56
Erhöhen des maximalen Parallelismuswerts. 56
Optimieren von Einfachdateien für Partitionierung. 57
    Optimieren von Einfachdateiquellen für die Partitionierung. 57
    Optimieren von Einfachdateizielen für die Partitionierung. 57
Optimieren von relationalen Datenbanken für Partitionierung. 58
    Optimieren der Quelldatenbank für Partitionierung. 58
    Optimieren der Zieldatenbank für Partitionierung. 58
Optimieren von Umwandlungen für Partitionierung. 59

Kapitel 7: Laufzeitoptimierung. 61

Laufzeitoptimierung - Übersicht. 61
Anwendungsdienstoptimierung. 61
    Analyst Service-Optimierung. 61
    Datenintegrationsdienst-Optimierung. 62
    Optimierung des Modell-Repository Service. 63
Überwachungsstatistiken. 63
Speicherzuordnung. 65
Datenobjekt-Caching. 66
    Datentypen für Cache-Tabellen. 67
    Cache-Optimierung für Datenobjekte. 69
Systemoptimierung. 69

Kapitel 8: SQL-Datendienstoptimierung. 71

SQL-Datendienstoptimierung - Übersicht. 71
Optimieren von Fremdanbieter-Client-Tools. 72
Optimierungsebenen des SQL-Datendiensts. 72
    Konfigurieren der SQL-Datendienst-Optimierungsebene für die Datenvorschau 73
    Konfigurieren der Optimierungsebene für bereitgestellte SQL Data Services. 73
    SQL-Datendienstabfrageplan. 74
    SQL-Abfrageplan anzeigen. 75
Eigenschaften des SQL-Datendiensts für Arbeitsspeicher und gleichzeitige Anfragen. 76
Ergebnissatz-Cache für einen SQL-Datendienst. 77
    Eigenschaften des Ergebnissatz-Caches für den SQL-Datendienst. 78
    Aktivieren des Ergebnissatz-Caches für einen SQL-Datendienst. 78
Dauerhaft virtuelle Daten in temporären Tabellen. 78
    Implementierung temporärer Tabellen. 79

Kapitel 9: Web-Dienst-Optimierung 80

Web-Dienst-Optimierung - Übersicht 80

Optimieren von HTTP-Anfragen 81

Komprimierung von Web-Dienst-Meldungen 81

Webdienst-Optimierungsebene 81

Konfigurieren der Webdienst-Optimierungsebene für die Datenvorschau 82

Konfigurieren der Optimierungsebene für bereitgestellte Web Services 82

Webdiensteigenschaften für Arbeitsspeicher und gleichzeitige Anfragen 83

Datenintegrationsdienst-Beispielkonfiguration für gleichzeitige Web-Dienst-Anfragen 85

Webdienst-Eigenschaft zum Konfigurieren einer aktiven DTM-Instanz 86

Zwischenspeichern von Web-Dienst-Ergebnissätzen 86

Aktivieren des Ergebnissatz-Caches für einen Web-Dienst 86

Verwaltung von Web-Dienst-Logs 87

Kapitel 10: Verbindungsoptimierung 88

Verbindungsoptimierung - Übersicht 88

Verbindungspooling 88

Poolingeigenschaften von Verbindungsobjekten 89

Größe von Datenbank-Netzwerkpaketen 90

Index 91

Einleitung

Das *Leistungsoptimierungs-Handbuch von Informatica* wurde für Administratoren und Entwickler geschrieben, die die Leistung von Data Services verbessern möchten. In diesem Handbuch wird vorausgesetzt, dass Sie mit den Betriebssystemen, Netzwerken, Client-Tools, Konzepten relationaler Datenbanken und mit Einfachdateien in Ihrer Umgebung vertraut sind. Informationen zur Optimierung der Leistung von Datenbanken, die in diesem Handbuch nicht behandelt wird, finden Sie in der Dokumentation Ihrer Datenbankprodukte.

Informatica-Ressourcen

Informatica-Netzwerk

Im Informatica-Netzwerk finden Sie den globalen Kundensupport von Informatica, die Informatica-Wissensdatenbank und andere Produktressourcen. Für den Zugriff auf das Informatica-Netzwerk besuchen Sie https://network.informatica.com.

Als Mitglied können Sie:

- zentral auf alle Ihre Informatica-Ressourcen zugreifen.
- Durchsuchen Sie die Wissensdatenbank nach Produktressourcen, einschließlich Dokumentation, häufig gestellter Fragen und bewährter Methoden.
- Zeigen Sie Informationen zur Produktverfügbarkeit an.
- Ihre Support-Fälle prüfen.
- Ihr lokales Informatica-Netzwerk für Benutzergruppen suchen und mit anderen Benutzern zusammenarbeiten.

Informatica-Wissensdatenbank

Verwenden Sie die Informatica-Wissensdatenbank, um das Informatica-Netzwerk nach Produktressourcen, wie z. B. Dokumentation, Ratgeberartikeln, bewährten Methoden und PAMs, zu durchsuchen.

Für den Zugriff auf die Wissensdatenbank besuchen Sie https://kb.informatica.com. Wenn Sie Fragen, Kommentare oder Ideen zur Wissensdatenbank haben, wenden Sie sich per E-Mail an das Team der Informatica-Wissensdatenbank unter KB_Feedback@informatica.com.

Informatica-Dokumentation

Navigieren Sie zur Informatica-Wissensdatenbank unter https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx, um die aktuelle Dokumentation für Ihr Produkt abzurufen.

Wenn Sie Fragen, Kommentare oder Ideen zu dieser Dokumentation haben, wenden Sie sich per E-Mail an das Informatica-Dokumentationsteam unter infa_documentation@informatica.com.

Informatica-Produktverfügbarkeitsmatrizen

Produktverfügbarkeitsmatrizen (PAMs) geben die Versionen der Betriebssysteme, Datenbanken und anderen Typen von Datenquellen und Zielen an, die in einer Produktversion unterstützt werden. Als Mitglied des Informatica-Netzwerks können Sie unter https://network.informatica.com/community/informatica-network/product-availability-matrices auf PAMs zugreifen.

Informatica Velocity

Bei Informatica Velocity handelt es sich um eine Sammlung von Tipps und bewährten Methoden, die von den professionellen Informatica-Diensten entwickelt wurden. Informatica Velocity basiert auf der Praxiserfahrung aus Hunderten von Datenmanagementprojekten und umfasst das kollektive Wissen unserer Berater, die mit Unternehmen aus der ganzen Welt an der Planung, Entwicklung, Bereitstellung und Wartung erfolgreicher Datenmanagementlösungen gearbeitet haben.

Als Mitglied des Informatica-Netzwerks können Sie unter http://velocity.informatica.com auf Informatica Velocity-Ressourcen zugreifen.

Wenn Sie Fragen, Anregungen oder Ideen zu Informatica Velocity haben, wenden Sie sich an die professionellen Informatica-Dienste unter ips@informatica.com.

Informatica Marketplace

Informatica Marketplace ist ein Forum, das Lösungen zur Erweiterung und Verbesserung Ihrer Informatica-Implementierungen bereitstellt. Indem Sie die zahlreichen Lösungen von Informatica-Entwicklern und -Partnern nutzen, können Sie Ihre Produktivität steigern und die Implementierungsdauer Ihrer Projekte verkürzen. Zugriff auf den Informatica Marketplace erhalten Sie unter https://marketplace.informatica.com.

Globaler Kundensupport von Informatica

Sie können sich telefonisch oder über den Online-Support mit einem globalen Support-Center im Informatica-Netzwerk in Verbindung setzen.

Die Telefonnummer des globalen Kundensupports von Informatica vor Ort finden Sie auf der Informatica-Website unter folgender Verknüpfung: http://www.informatica.com/us/services-and-training/support-services/global-support-centers.

Als Mitglied des Informatica-Netzwerks können Sie den Online-Support unter http://network.informatica.com verwenden.

KAPITEL 1

Leistungsoptimierung - Übersicht

Dieses Kapitel umfasst die folgenden Themen:

- Leistungsoptimierung - Übersicht, 10
- Target-Engpässe, 11
- Quellengpässe, 11
- Mapping-Engpässe, 12
- Engpässe im Computersystem, 12
- Laufzeit-Engpässe, 13
- Engpässe bei der SQL-Datendienstoptimierung, 13
- Engpässe bei der Web-Dienst-Optimierung, 14
- Verbindungsengpässe, 14

Leistungsoptimierung - Übersicht

Die Leistungsoptimierung soll Leistungsengpässe beseitigen. Ein Engpass ist ein Bereich im Mapping, der am häufigsten ausgeführt wird und den geringsten Durchsatz aufweist. Ein Engpass reduziert die Gesamtleistung des Mappings.

Um ein Mapping zu optimieren, identifizieren Sie einen Leistungsengpass, beseitigen Sie ihn und identifizieren Sie dann den nächsten Leistungsengpass. Optimieren Sie immer nur eine Mapping-Komponente gleichzeitig. Sie können die Ausführungsdauer eines Mappings vor und nach der Änderung messen, um zu überprüfen, ob sich die Optimierung auf die Leistung auswirkt.

Sie können die Mapping-Komponenten in der folgenden Reihenfolge optimieren:

1. Targets
2. Quellen
3. Mappings
4. Umwandlungen
5. Informatica-Umgebung im Administrator Tool
6. Das Computersystem
7. Datendienst oder Webdienst

Identifizieren Sie Leistungsengpässe mit den folgenden Methoden:

- Führen Sie Test-Mappings aus. Sie können ein Test-Mapping zum Lesen aus einer Einfachdateiquelle oder zum Schreiben in ein Einfachdateiziel konfigurieren, um Quell- und Target-Engpässe zu bestimmen.
- Analysieren Sie Leistungsdetails. Analysieren Sie Leistungsdetails, z. B. Optimierungsmethoden, um zu bestimmen, wo die Mapping-Leistung abnimmt.
- Überwachen Sie die Systemleistung. Mit Systemüberwachungstools können Sie Prozentwerte für CPU-Auslastung, E/A-Wartezeiten, Auslagerung und Systemressourcen-Nutzung anzeigen.

Target-Engpässe

Ein Target-Engpass ist ein Leistungsabfall, wenn der Data Integration Service in ein Target schreibt. Target-Engpässe können auftreten, wenn die Datenbank kurze Prüfpunktintervalle oder kleine Datenbank-Netzwerkpakete verwendet.

Der häufigste Leistungsengpass tritt auf, wenn der Data Integration Service in eine Target-Datenbank schreibt. Wenn die Datenbank kurze Prüfpunktintervalle verwendet, wird die Datenbankverarbeitung häufiger beim Schreiben eines Prüfpunkts verlangsamt. Kleine Datenbank-Netzwerkpakete können Engpässe verursachen. Sie können zulassen, dass größere Datenpakete das Netzwerk gleichzeitig durchlaufen.

Zum Bestimmen eines Target-Engpasses können Sie eine Kopie des Mappings erstellen, die statt eines Datenbank-Targets über ein Einfachdatei-Target verfügt. Wenn sich die Leistung erheblich verbessert, liegt ein Target-Engpass vor. Wenn das Mapping bereits in ein Einfachdatei-Target schreibt, liegt wahrscheinlich kein Target-Engpass vor.

Quellengpässe

Ein Quellengpass ist ein Leistungsabfall beim Lesen aus einer Quelldatenbank durch den Data Integration Service. Quellengpässe können auftreten, wenn die Quellabfrage nicht effizient ist oder wenn die Datenbank-Netzwerkpakete klein sind.

Wenn das Mapping aus einer relationalen Quelle liest, können Sie mit den folgenden Methoden Quellengpässe bestimmen:

- Fügen Sie dem Mapping eine Filter-Umwandlung hinzu. Fügen Sie die Filter-Umwandlung nach der Quelle hinzu. Legen Sie die Filterbedingung auf false fest, damit die Filter-Umwandlung keine Daten zurückgibt. Wenn sich die Ausführungsdauer des Mappings nicht erheblich geändert hat, liegt ein Quellengpass im Mapping vor.
- Erstellen Sie ein Lese-Test-Mapping. Erstellen Sie eine Kopie des Mappings, entfernen Sie jedoch alle Umwandlungen, Joins und Abfragen. Verbinden Sie die Quelle mit einem Target. Wenn die Mapping-Leistung etwa so hoch wie die Leistung des ursprünglichen Mappings ist, liegt ein Quellengpass vor.
- Führen Sie die Lese-Abfrage direkt für die Quelldatebank aus. Kopieren Sie die Lese-Abfrage aus dem Mapping-Log. Führen Sie die Abfrage mit einem Abfragetool, z. B. isql, für die Quelldatenbank aus. Messen Sie die Laufzeit und die Zeit, bis die Abfrage eine Zeile zurückgibt.

Mapping-Engpässe

Wenn Sie feststellen, dass kein Quell- oder Target-Engpass vorliegt, besteht möglicherweise ein Mapping-Engpass. Mapping-Engpässe können durch geringe Cache-Größe, geringen Pufferspeicher und kurze Commit-Intervalle verursacht werden.

Um einen Mapping-Engpass zu bestimmen, analysieren Sie die Leistungsdetails im Mapping-Log. Zu den Leistungsdetails zählen Informationen über jede Umwandlung, z. B. die Anzahl von Eingabezeilen, Ausgabezeilen und Fehlerzeilen.

Sie können auch vor jeder Target-Definition eine Filter-Umwandlung hinzufügen. Legen Sie die Filterbedingung auf false fest, damit die Filter-Umwandlung keine Daten in die Target-Tabellen lädt. Wenn das Ausführen des neuen Mappings genauso lange wie das Ausführen des ursprünglichen Mappings dauert, besteht ein Mapping-Engpass.

Engpässe im Computersystem

Wenn Sie Informatica Services unter Windows oder UNIX ausführen, können Sie die Ressourcennutzung anzeigen. Verwenden Sie unter Windows den Task Manager. UNIX bietet mehrere Tools, mit denen Sie die Leistung überprüfen können.

Bestimmen von Systemengpässen unter Windows

Sie können im Task-Manager auf den Registerkarten „Leistung“ und „Prozesse“ Systeminformationen anzeigen. Die Registerkarte „Leistung“ im Task-Manager bietet eine Übersicht über die CPU-Auslastung und die Summe des genutzten Arbeitsspeichers. Mit dem Systemmonitor können Sie ausführlichere Informationen anzeigen.

In der folgenden Tabelle werden die Systeminformationen beschrieben, mit denen Sie im Systemmonitor von Windows ein Diagramm erstellen können:

Eigenschaft	Beschreibung
Prozessorzeit (%)	Wenn mehrere CPUs enthalten sind, können Sie die prozentuale Prozessorzeit jeder CPU überwachen.
Seiten/s	Wenn der Wert von „Seiten/s“ höher als 5 ist, wird möglicherweise der Arbeitsspeicher übermäßig beansprucht (Thrashing).
Physischer Datenträger: Zeit (%)	Der Prozentsatz der Zeit, während der der physische Datenträger Lese- oder Schreibanfragen ausführt.
Physischer Datenträger: Aktuelle Warteschlangenlänge	Die Anzahl der Benutzer, die auf Zugriff auf denselben Datenträger warten.
Server: Bytes pro Sekunde	Die Anzahl von Bytes, die der Server an das Netzwerk gesendet und vom Netzwerk empfangen hat.

Bestimmen von Systemengpässen unter UNIX

Verwenden Sie zum Bestimmen von Systemengpässen unter UNIX die folgenden Tools:

- top. Zeigt die gesamte Systemleistung an. Dieses Tool zeigt die Systemauslastung, Speicherauslastung und Swap-Auslastung für das System und für einzelne auf dem System ausgeführte Prozesse an.
- iostat. Überwacht den Ladevorgang für jeden Datenträger, der mit dem Datenbankserver verbunden ist. iostat zeigt den Prozentsatz der Zeit an, während derer der Datenträger physisch aktiv ist. Wen Sie Datenträgerarrays verwenden, nutzen Sie statt iostat die mit den Datenträgerarrays bereitgestellten Dienstprogramme.
- vmstat. Überwacht Datenträgerauslagerungs-Aktionen.
- sar. Zeigt ausführliche Systemaktivitätsberichte zu CPU-, Speicher- und Datenträgerauslastung an. Mit diesem Tool können Sie CPU-Ladevorgänge überwachen. Es liefert prozentuale Nutzungswerte zu Benutzer, System, Leerlaufzeit und Wartezeit. Mit diesem Tool können Sie auch Datenträgerauslagerungs-Aktionen überwachen.

Laufzeit-Engpässe

Aktivieren Sie Leistungsfunktionen und passen Sie Datenintegrationsdienst-Eigenschaften an, um die Leistung von Mappings zu optimieren. Konfigurieren Sie im Administrator Tool die Optimierungseinstellungen für den Datenintegrationsdienst und den Modellrepository-Dienst.

Reservieren Sie Speicher, um eine optimale Systemleistung zu erzielen, und konfigurieren Sie Fehlertracingebenen, um die Anzahl der Log-Ereignisse zu verringern, die der Datenintegrationsdienst beim Ausführen des Mappings generiert.

Sie können die maximale Speichermenge konfigurieren, die der Datenintegrationsdienst für das Ausführen gleichzeitiger Anfragen zuteilt. Sie können auch die maximale Speichermenge begrenzen, die der Datenintegrationsdienst für Anfragen zuteilt.

Sie können den Ergebnissatz-Cache konfigurieren, damit der Datenintegrationsdienst die Ergebnisse des DTM-Prozesses, der mit den einzelnen SQL-Datendienst-Abfragen und Webdienst-Anfragen verknüpft ist, zwischenspeichern kann.

Engpässe bei der SQL-Datendienstoptimierung

Sie können SQL-Datendienste optimieren, um die Leistung zu verbessern, wenn Endbenutzer mit Fremdanbieter-Client-Tools SQL-Abfragen dieser Datendienste ausführen. Wenn ein SQL-Datendienst ein virtuelles Tabellen-Mapping verwendet, können Sie die Umwandlungen und das Mapping optimieren.

Sie können den JDBC-Treiber optimieren, um die Leistung beim Abfragen eines SQL-Datendienstes zu verbessern. Sie können auch den Datenobjekt-Cache für den Data Integration Service zum Verbessern der Leistung von Mappings und SQL-Abfragen konfigurieren.

Engpässe bei der Web-Dienst-Optimierung

Sie können Web-Dienste optimieren, um die Leistung beim Ausführen von Web-Dienst-Anfragen durch den Data Integration Service zu verbessern. Passen Sie den Data Integration Service an, um Arbeitsspeicher zu verwalten, gleichzeitige Web-Dienst-Anfragen zu behandeln und einen DTM-Prozess aktiv zu lassen,damit er mehrere Web-Dienst-Anfragen verarbeiten kann.

Um die Web-Dienst-Leistung zu verbessern, komprimieren Sie Web-Dienst-Meldungen, optimieren Sie HTTP-Anfragen und konfigurieren Sie den Datenobjekt-Cache.

Verbindungsengpässe

Sie können Verbindungen optimieren, um die Leistung zu verbessern. Sie können den Pool inaktiver Verbindungsinstanzen für eine Datenbankverbindung verwalten. Sie können die Netzwerkpaketgröße erhöhen, damit größere Datenpakete das Netzwerk gleichzeitig durchlaufen können.

KAPITEL 2

Target-Optimierung

Dieses Kapitel umfasst die folgenden Themen:

- Ziel-Optimierung - Übersicht, 15
- Optimierung von Einfachdateizielen, 15
- Datenbankprüfpunkt-Intervalle, 16
- Massenladevorgänge, 16
- Datenbank-Ziel-Optimierung, 17

Ziel-Optimierung - Übersicht

Optimieren Sie Targets, damit der Datenintegrationsdienst effizient in Targets schreiben kann. Sie können vor dem Ausführen eines Mappings Indizes und Schlüsselbeschränkungen löschen, die Anzahl der Prüfpunktintervalle in der Datenbank erhöhen, in den Schreibeigenschaften für ein Datenbankobjekt Massenladen konfigurieren und eine Oracle-Datenbank optimieren.

Verwenden Sie zum Optimieren des Targets die folgenden Optimierungsverfahren:

- Optimieren Sie Einfachdateiziele.
- Vergrößern Sie Datenbankprüfpunkt-Intervalle.
- Verwenden Sie Massenladevorgänge.
- Optimieren Sie Oracle-Datenbanken.

Optimierung von Einfachdateizielen

Sie können durch Optimieren von Einfachdateizielen die Mapping-Leistung verbessern. Sie können auch zum Verbessern der Leistung Umwandlungsaufgaben an einen Befehl übertragen.

Zum Beheben von Engpässen bei Einfachdateizielen können Sie die folgenden Lösungen verwenden:
Übertragen Sie Umwandlungsaufgaben an einen Befehl statt an den Datenintegrationsdienst.

Sie können die Mapping-Leistung verbessern, indem Sie Umwandlungsaufgaben an einen Befehl statt an den Datenintegrationsdienst übertragen. Sie können auch mit einem Befehl Target-Daten sortieren oder komprimieren. Konfigurieren Sie im Developer Tool die Befehlseigenschaft in den Laufzeiteigenschaften für ein Einfachdateiziel.

Verwenden Sie unter UNIX einen beliebigen gültigen UNIX-Befehl oder ein beliebiges Shell-Skript. Verwenden Sie unter Windows einen beliebigen gültigen DOS-Befehl oder eine Batchdatei. Der Einfachdatei-Writer sendet die Daten an den Befehl und nicht an ein Einfachdateiziel.

Verwenden Sie beispielsweise den folgenden Befehl, um aus den Target-Daten eine komprimierte Datei zu generieren:

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

Schreiben Sie in ein Einfachdateiziel, das lokal auf dem Dienstprozess-Knoten vorhanden ist.

Wenn der Datenintegrationsdienst auf einem einzelnen Knoten ausgeführt wird und in ein Einfachdateiziel schreibt, können Sie die Mapping-Leistung optimieren, indem Sie in ein lokales Einfachdateiziel auf dem Dienstprozess-Knoten schreiben.

Datenbankprüfpunkt-Intervalle

Die Leistung des Data Integration Service fällt immer ab, wenn dieser auf das Ausführen eines Prüfpunkts durch die Datenbank wartet.

Zum Beheben von Datenbankprüfpunkt-Engpässen können Sie die folgende Lösung verwenden:
Vergrößern Sie das Prüfpunkt-Intervall in der Datenbank.

Um die Anzahl der Prüfpunkte zu verringern und die Leistung zu erhöhen, vergrößern Sie das Prüfpunktintervall in der Datenbank.

Durch das Verringern der Anzahl von Prüfpunkten verbessern Sie zwar die Leistung, jedoch erhöht sich auch die Wiederherstellungszeit, wenn die Ausführung der Datenbank unerwartet beendet wird.

Massenladevorgänge

Wenn Sie Massenladen verwenden, umgeht der Data Integration Service das Datenbank-Log, wodurch sich die Leistung erhöht.

Zum Beheben von Engpässen beim Massenladen können Sie die folgenden Lösungen verwenden:
Konfigurieren Sie in den Schreibeigenschaften für ein Datenobjekt Massenladen.

Mit Massenladen können Sie die Leistung eines Mappings erhöhen, das eine große Menge von Daten in eine DB2-, Sybase ASE-, Oracle- oder Microsoft SQL Server-Datenbank einfügt.

Die Target-Datenbank kann kein Rollback ausführen, wenn nicht in das Datenbank-Log geschrieben wird. Dies kann dazu führen, dass keine Wiederherstellung möglich ist. Wenn Sie Massenladen verwenden, wägen Sie die Vorteile einer höheren Mapping-Leistung gegen die Fähigkeit zum Wiederherstellen eines unvollständigen Mappings ab.

Datenbank-Ziel-Optimierung

Sie können die Zieldatenbank optimieren, indem Sie die Storage-Klausel, Speicherplatzzuweisung sowie das Rollback- und das Undo-Segment überprüfen.

Zum Beheben von Engpässen bei Zieldatenbanken können Sie die folgenden Lösungen verwenden:

Stellen Sie sicher, dass die Datenbank Rollback- oder Undo-Segmente in geeigneten Tabellenbereichen, nach Möglichkeit auf unterschiedlichen Datenträgern, speichert.

Beim Schreiben in Datenbanken verwendet die Datenbank während des Ladens Rollback- oder Undo-Segmente. Bitten Sie den Datenbankadministrator, sicherzustellen, dass die Datenbank Rollback- oder Undo-Segmente in geeigneten Tabellenbereichen, nach Möglichkeit auf unterschiedlichen Datenträgern, speichert. Die ROLLBACK- oder UNDO-Segmente sollten außerdem über entsprechende STORAGE-Klauseln verfügen.

Passen Sie das Wiederholen-Protokoll der Datenbank an.

Passen Sie zum Optimieren der Datenbank das Datenbank-Wiederholen-Protokoll an. Die Datenbank protokolliert Ladevorgänge im Wiederholen-Protokoll. Stellen Sie eine optimale Größe von Wiederholen-Protokoll und Puffer sicher. Sie können für die Oracle-Datenbank die Wiederholen-Protokoll-Eigenschaften in der Datei init.ora anzeigen.

Stellen Sie mit dem IPC-Protokoll eine Verbindung zu einer Oracle-Datenbank her.

Wenn der Datenintegrationsdienst auf einem einzelnen Knoten ausgeführt wird und die Oracle-Instanz lokal auf dem Dienstprozessknoten vorhanden ist, können Sie für die Verbindung mit der Oracle-Datenbank das IPC-Protokoll verwenden, um die Leistung zu optimieren. Sie können die Oracle-Datenbankverbindung in listener.ora und tnsnames.ora festlegen.

KAPITEL 3

Quelloptimierung

Dieses Kapitel umfasst die folgenden Themen:

- Quelloptimierung - Übersicht, 18
- Optimierung von Einfachdateiquellen, 19
- Abfrageoptimierung, 19
- Bedingungsfilter, 20
- Select Distinct, 20
- Hinweise, 21
- Einschränkungen, 22
- Optimierung von konfigurierten Datenobjekten, 23
- Datenbankquelloptimierung, 24

Quelloptimierung - Übersicht

Optimieren Sie Einfachdateiquellen, relationale und benutzerdefinierte Dateiquellen, damit der Data Integration Service Quelldaten effizient lesen kann.

Verwenden Sie zum Optimieren von Quellen die folgenden Optimierungsverfahren:

- Lesen Sie Quelldaten effizient.
- Verwenden Sie Abfrageoptimierungsverfahren.
- Verwenden Sie Bedingungsfilter für die SQL-Abfrage.
- Wählen Sie eindeutige Werte aus der Quelle aus.
- Wenden Sie Hinweise auf die SQL-Abfrage an.
- Konfigurieren Sie Einschränkungen für logische Datenobjekte, physische Datenobjekte und virtuelle Tabellen.
- Legen Sie konfigurierte Datenobjekte für die Optimierung fest.
- Konfigurieren Sie Oracle-, Sybase- und Microsoft SQL Server-Datenbanken für die Optimierung.

Optimierung von Einfachdateiquellen

Konfigurieren Sie die Eigenschaften für Einfachdateiquellen, damit der Data Integration Service Quelldaten effizient lesen kann.

Zum Beheben von Engpässen bei Einfachdateiquellen können Sie die folgenden Lösungen verwenden:

Verwenden Sie in den Formateigenschaften einer durch Trennzeichen getrennten Einfachdatei keine Anführungszeichen oder Escape-Zeichen.

Wenn Sie ein Escape-Zeichen angeben, liest der Data Integration Service das Delimiterzeichen als reguläres Zeichen in der Zeichenfolge. Sie können die Mapping-Leistung etwas verbessern, wenn die Quelldatei keine Anführungszeichen und Escape-Zeichen enthält.

Legen Sie die Anzahl von Bytes fest, die der Data Integration Service pro Zeile liest.

Wenn das Mapping aus einer Einfachdateiquelle liest, können Sie die Mapping-Leistung erhöhen, indem Sie die Anzahl der vom Data Integration Service pro Zeile gelesenen Bytes festlegen. Konfigurieren Sie in den Laufzeiteigenschaften für Einfachdateiquellen die Länge des zeilensequentiellen Puffers.

Der Data Integration Service liest standardmäßig 1024 Bytes pro Zeile. Wenn jede Zeile in der Quelldatei kürzer als die Standardeinstellung ist, können Sie in den Mapping-Eigenschaften die Länge des zeilensequentiellen Puffers verringern.

Abfrageoptimierung

Wenn ein Mapping mehrere Quelltabellen in einem konfigurierten Datenobjekt verknüpft, können Sie möglicherweise die Leistung verbessern, indem Sie die Abfrage mit Optimierungshinweisen optimieren. Zudem kann eine einzelne SELECT-Anweisung für eine Tabelle mit einer ORDER BY- oder GROUP BY-Klausel von einer Optimierung wie z. B. dem Hinzufügen von Indizes profitieren.

Zum Beheben von Abfrageengpässen können Sie die folgenden Lösungen verwenden:

Erstellen Sie Optimierungshinweise, anhand derer die Datenbank erkennt, wie die Abfrage für einen bestimmten Satz von Quelltabellen ausgeführt werden soll.

In der Regel bestimmt das Datenbankoptimierungsprogramm die effizienteste Methode zum Verarbeiten der Quelldaten. Jedoch kennen Sie möglicherweise Eigenschaften der Quelltabellen, die das Datenbankoptimierungsprogramm nicht kennt. Der Datenbankadministrator kann Optimierungshinweise erstellen, anhand derer die Datenbank erkennt, wie die Abfrage für einen bestimmten Satz von Quelltabellen ausgeführt werden soll.

Konfigurieren Sie Optimierungshinweise, damit Zeilen so schnell wie möglich zurückgegeben werden, statt alle Zeilen auf einmal zurückzugeben.

Verwenden Sie Optimierungshinweise, wenn zwischen dem Beginn der Abfrageausführung und dem Empfang der ersten Zeile von Daten durch den Datenintegrationsdienst ein langer Zeitraum liegt. Konfigurieren Sie Optimierungshinweise, damit Zeilen so schnell wie möglich zurückgegeben werden, statt alle Zeilen auf einmal zurückzugeben. So kann der Datenintegrationsdienst Zeilen gleichzeitig mit der Ausführung der Abfrage verarbeiten.

Erstellen Sie einen Index für die ORDER BY- oder GROUP BY-Spalten.

Abfragen mit ORDER BY- oder GROUP BY-Klauseln können vom Erstellen eines Index für die ORDER BY- bzw. GROUP BY-Spalten profitieren. Nachdem Sie die Abfrage optimiert haben, verwenden Sie die SQL-Überschreibungsoption, um die Vorteile dieser Änderungen umfassend zu nutzen.

Konfigurieren Sie die Datenbank für die Ausführung paralleler Abfragen.

Sie können auch die Quelldatenbank für die Ausführung paralleler Abfragen konfigurieren, um die Leistung zu verbessern. Weitere Informationen zum Konfigurieren von parallelen Abfragen finden Sie in der Datenbankdokumentation.

Die Abfrage, die der Datenintegrationsdienst zum Lesen der Daten verwendet, wird in der virtuellen Datenbank eines SQL-Datendienstes angezeigt. Sie finden die Abfrage auch im konfigurierten Datenobjekt. Lassen Sie den Datenbankadministrator die Abfrage analysieren und erstellen Sie dann Optimierungshinweise und Indizes für die Quelltabellen.

Bedingungsfilter

Ein einfacher Quellfilter für die Quelldatenbank kann manchmal die Leistung beeinträchtigen, da keine Indizes vorhanden sind. Mit dem Bedingungsfilter im konfigurierten Datenobjekt können Sie die Leistung verbessern.

Zum Beheben von Bedingungsfilter-Engpässen können Sie die folgende Lösung verwenden:
Verwenden Sie den Bedingungsfilter für mehrere Mappings, die gleichzeitig aus derselben Quelle lesen.

Wenn mehrere Mappings gleichzeitig aus derselben Quelle lesen, wird die Leistung möglicherweise durch den Bedingungsfilter verbessert.

Die Leistung einiger Mappings ist jedoch möglicherweise höher, wenn Sie die Quelldaten in der Quelldatenbank filtern. Sie können das Mapping sowohl mit dem Datenbankfilter als auch mit dem Bedingungsfilter testen, um zu bestimmen, welche Methode die Leistung verbessert.

Select Distinct

Mit der Select Distinct-Option können Sie eindeutige Werte aus Quellen in einem konfigurierten Datenobjekt auswählen. Wenn Sie Select Distinct verwenden, fügt der Data Integration Service der Standard-SQL-Abfrage eine SELECT DISTINCT-Anweisung hinzu.

Zum Beheben von Select Distinct-Engpässen können Sie die folgende Lösung verwenden:
Verwenden Sie die Select Distinct-Option, um unnötige Daten zu einem früheren Zeitpunkt im Datenfluss zu filtern.

Verwenden Sie die Select Distinct-Option für das konfigurierte Datenobjekt, wenn der Data Integration Service eindeutige Werte aus einer Quelle auswählen soll. Verwenden Sie die Select Distinct-Option, um unnötige Daten zu einem früheren Zeitpunkt im Datenfluss zu filtern. Dies kann die Leistung verbessern.

Beispielsweise können Sie mit der Select Distinct-Option eindeutige Kunden-IDs aus einer Tabelle extrahieren, in der Umsatzsummen aufgeführt werden. Wenn Sie in einem Mapping das konfigurierte Datenobjekt verwenden, filtert der Data Integration Service unnötige Daten zu einem früheren Zeitpunkt im Datenfluss heraus. Dies kann die Leistung verbessern.

Hinweise

Sie können der Quell-SQL-Abfrage Hinweise hinzufügen, um Anweisungen an ein Datenbankoptimierungsprogramm zu übergeben. Das Optimierungsprogramm nutzt die Hinweise, um einen Abfrageausführungsplan für den Zugriff auf die Quelle auszuwählen.

Das Feld „Hinweise“ wird in der Ansicht **Abfrage** einer relationalen Datenobjektinstanz oder eines konfigurierten Datenobjekts angezeigt. Die Quelldatenbank muss eine Oracle-, Sybase-, IBM DB2- oder Microsoft SQL Server-Datenbank sein.. Für andere Datenbanktypen wird das Feld „Hinweise“ nicht angezeigt.

Wenn der Data Integration Service die Quellabfrage generiert, fügt er der Abfrage die SQL-Hinweise genau so hinzu, wie Sie sie im Developer Tool eingegeben haben. Der Data Integration Service parst die Hinweise nicht. Wenn Sie das Mapping ausführen, das die Quelle enthält, wird die Abfrage mit den Hinweisen in der Abfrage im Mapping-Log angezeigt.

Der Data Integration Service fügt die SQL-Hinweise an einer vom Datenbanktyp abhängigen Position in der Abfrage ein. Informationen zur Syntax für Hinweise finden Sie in der Datenbankdokumentation.

Oracle

Der Data Integration Service fügt Hinweise direkt nach dem Schlüsselwort SELECT/UPDATE/INSERT/DELETE ein.

```
SELECT /*+ <Hinweise> */ FROM …
```

Das Zeichen „+“ gibt den Anfang der Hinweise an.

Die Hinweise sind in einem Kommentar (/* ... */ oder --... bis Zeilenende) enthalten

Sybase

Der Data Integration Service fügt Hinweise nach der Abfrage ein. Konfigurieren Sie einen Plannamen im Hinweis.

```
SELECT … PLAN <Plan>
```

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

IBM DB2

Sie können die OPTIMIZE FOR-Klausel als Hinweis eingeben. Der Data Integration Service fügt die Klausel am Ende der Abfrage ein.

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

Durch die OPTIMIZE FOR-Klausel erkennt das Datenbankoptimierungsprogramm, wie viele Zeilen die Abfrage möglicherweise verarbeitet. Die Klausel beschränkt nicht die Anzahl der Zeilen. Wenn die Datenbank mehr als <n> Zeilen verarbeitet, nimmt die Leistung möglicherweise ab.

Microsoft SQL Server

Der Data Integration Service fügt Hinweise am Ende der Abfrage als OPTION-Klausel hinzu.

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

Regeln und Richtlinien für Hinweise

Beachten Sie beim Konfigurieren von Hinweisen für SQL-Abfragen die folgenden Regeln und Richtlinien:

- Wenn Sie Pushdown-Optimierung aktivieren oder in einem relationalen Datenobjekt einen Semi-Join verwenden, ändert sich die ursprüngliche Quellabfrage. Der Data Integration Service wendet keine Hinweise auf die geänderte Abfrage an.
- Sie können Hinweise mit Join- und Filterüberschreibungen kombinieren. Wenn Sie jedoch eine SQL-Überschreibung konfigurieren, hat diese Vorrang und der Data Integration Service wendet die anderen Überschreibungen nicht an.
- Die Ansicht **Abfrage** ist eine einfache oder eine erweiterte Ansicht. Wenn Sie in der einfachen Ansicht einen Hinweis mit einer Filter-, Sortierungs- oder Join-Überschreibung eingeben, zeigt das Developer Tool in der erweiterten Ansicht die vollständige Abfrageüberschreibung an.

Erstellen von Hinweisen

Erstellen Sie Hinweise, die als Anweisungen zum Bestimmen eines Abfrageplans an das Datenbankoptimierungsprogramm gesendet werden sollen.

1. Öffnen Sie das konfigurierte Datenobjekt oder die Instanz des relationalen Datenobjekts.
2. Wählen Sie die Ansicht **Lesen** aus.
3. Wählen Sie die Ausgabeumwandlung aus.
4. Wählen Sie die **Abfrage**-Eigenschaften aus.
5. Wählen Sie die einfache Abfrage aus.
6. Klicken Sie neben dem Feld **Hinweise** auf **Bearbeiten**.

   Das Dialogfeld **Hinweise** wird angezeigt.
7. Geben Sie im Feld **SQL-Abfrage** den Hinweis ein.

   Der Hinweis wird nicht vom Developer Tool validiert.
8. Klicken Sie auf **OK**.
9. Speichern Sie das Datenobjekt.

Einschränkungen

Der Datenintegrationsdienst kann Beschränkungen aus relationalen Quellen, Einfachdateiquellen, logischen Datenobjekten oder virtuellen Tabellen lesen. Eine Einschränkung ist ein Bedingungsausdruck, dem die Werte in einer Datenzeile entsprechen müssen.

Wenn der Datenintegrationsdienst Einschränkungen liest, lässt er basierend auf der angewendeten Optimierungsmethode unter Umständen die Zeilen außer Acht, die nicht mit TRUE bewertet werden.

Vor dem Einrichten einer Einschränkung müssen Sie sicherstellen, dass die Quelldaten der von der Einschränkung festgelegten Bedingung entsprechen.

Beispiel: Eine Quelldatenbank verfügt über eine ALTER-Spalte, die zwei Zeilen mit ALTER < 70 aufweist. Sie können eine Einschränkung mit ALTER < 70 in der Quelldatenbank einrichten. Der Datenintegrationsdienst liest Datensätze aus der Quelldatenbank mit der Einschränkung ALTER < 70. Wenn der Datenintegrationsdienst Datensätze mit ALTER >= 70 liest, lässt er die Zeilen mit ALTER >= 70 unter Umständen außer Acht.

In der Datenbank können Sie SQL-Befehle verwenden, um Einschränkungen für die Datenbankumgebung einzurichten, wenn Sie eine Verbindung zur Datenbank herstellen. Der Datenintegrationsdienst führt die Verbindungsumgebungs-SQL bei jeder Verbindung mit der Datenbank aus.

Verwenden Sie das Developer Tool, um Einschränkungen für logische Datenobjekte, physische Datenobjekte und virtuelle Tabellen einzurichten. Wenn Sie eine Einschränkung einrichten, müssen Sie einen Ausdruck eingeben, der für jede Datenzeile mit TRUE bewertet wird.

Konfigurieren von Einschränkungen

Sie können Einschränkungen zu relationalen Datenobjekten, Einfachdatei-Datenobjekten, benutzerdefinierten Datenobjekten, logischen Datenobjekten und virtuellen Tabellen hinzufügen. Nach dem Hinzufügen einer Einschränkung können Sie diese bearbeiten oder löschen.

1. Öffnen Sie über die Ansicht **Objekt-Explorer** das Mapping, das das relationale Datenobjekt enthält, welches als Leseumwandlung hinzugefügt wurde. Oder öffnen Sie das Einfachdatei-Datenobjekt, das benutzerdefinierte Datenobjekt, das logische Datenobjekt oder die virtuelle Tabelle.
   - Wählen Sie zum Einrichten von Einschränkungen für ein als Leseumwandlung hinzugefügtes relationales Datenobjekt die Leseumwandlung im Mapping aus. Wählen Sie in der Ansicht **Eigenschaften** die Registerkarte **Erweitert** aus.
   - Wählen Sie zum Einrichten von Einschränkungen für ein Einfachdatei-Datenobjekt die Ansicht **Erweitert** aus und erweitern Sie den Abschnitt **Laufzeit: Lesen**.
   - Wählen Sie zum Einrichten von Einschränkungen für ein benutzerdefiniertes Datenobjekt die Ansicht **Lesen** und den **Ausgabe**port der Quellumwandlung aus. Wählen Sie in der Ansicht **Eigenschaften** die Registerkarte **Erweitert** aus.
   - Um Einschränkungen für ein logisches Datenobjekt einzurichten, wählen Sie ein logisches Datenmodell und das logische Datenobjekt aus. Wählen Sie in der Ansicht **Eigenschaften** die Registerkarte **Erweitert** aus.
   - Um Einschränkungen für eine virtuelle Tabelle einzurichten, öffnen Sie die virtuelle Tabelle über den SQL-Endpunkt. Wählen Sie in der Ansicht **Eigenschaften** die Registerkarte **Erweitert** aus.
2. Klicken Sie auf das Feld „Wert“ für Einschränkungen.

   Das Dialogfeld **Einschränkungen** wird angezeigt.
3. Klicken Sie auf **Neu**, um den Ausdruckseditor zu öffnen.
4. Konfigurieren Sie die Einschränkungslogik und verwenden Sie Ausdrucksfunktionen und -spalten als Parameter.
5. Klicken Sie auf **Validieren**.
6. Klicken Sie auf **OK**.

Optimierung von konfigurierten Datenobjekten

Sie können konfigurierte Datenobjekte zum Verbessern der Leistung konfigurieren. Sie können in einem konfigurierten Datenobjekt die SQL-Abfrage optimieren, Bedingungsfilter verwenden und individuelle Werte aus der Quelle auswählen.

Zum Beheben von Engpässen bei konfigurierten Datenobjekten können Sie die folgenden Lösungen verwenden:

Erstellen Sie eine benutzerdefinierte Abfrage, um eine besondere SELECT-Anweisung zum Lesen von Quelldaten durch den Data Integration Service auszugeben.

Die benutzerdefinierte Abfrage ersetzt die Standardabfrage, die der Data Integration Service zum Lesen von Daten aus Quellen verwendet.

Filtern Sie Zeilen, wenn der Data Integration Service Quelldaten liest.

Wenn Sie eine Filterbedingung einschließen, fügt der Data Integration Service der Standardabfrage eine WHERE-Klausel hinzu.

Wählen Sie individuelle Werte aus der Quelle aus.

Wenn Sie Select Distinct wählen, fügt der Data Integration Service der Standard-SQL-Abfrage eine SELECT DISTINCT-Klausel hinzu.

Wenden Sie Datenbankhinweise an.

Sie können der Quell-SQL-Abfrage Hinweise hinzufügen, um Anweisungen an ein Datenbankoptimierungsprogramm zu übergeben.

Konfigurieren Sie Einschränkungen für Quelldaten.

Wenn Sie Einschränkungen für Einfachdateien und relationale Tabellen in einem benutzerdefinierten Datenobjekt konfigurieren, lässt der Datenintegrationsdienst die Zeilen außer Acht, die nicht mit TRUE bewertet werden.

Datenbankquelloptimierung

Wenn die Quelldatenbank eine Oracle-Datenbank ist, können Sie die Datenintegrationsdienst-Leistung optimieren, indem Sie zum Herstellen der Verbindung mit der Oracle-Datenbank das IPC-Protokoll verwenden. Sie können die temporäre Datenbank auf einen Datenträgerarray verschieben, um die Leistung zu verbessern.

Zum Beheben von Engpässen bei Datenbankquellen können Sie die folgenden Lösungen verwenden:

Verwenden Sie zum Herstellen der Verbindung mit der Oracle-Datenbank das IPC-Protokoll.

Wenn der Datenintegrationsdienst auf einem einzelnen Knoten ausgeführt wird und die Oracle-Instanz lokal auf dem Dienstprozessknoten ausgeführt wird, können Sie für die Verbindung mit der Oracle-Datenbank das IPC-Protokoll verwenden, um die Leistung zu optimieren. Sie können die Oracle-Datenbankverbindung in listener.ora und tnsnames.ora festlegen.

Verschieben Sie die temporäre Datenbank und das Wiederholen-Protokoll auf Datenträgerarrays oder schnellere Treiber.

Wenn Sie große Tabellen auf einer Datenbank zusammenfügen, können Sie einen Array unabhängiger Datenträger (RAID) für den Cache-Speicherort verwenden. Alternativ können Sie auch mehr Dateien zur primären Dateiengruppe auf anderen Datenträgern hinzufügen, um die Datenmenge auf die Datenträger aufzuteilen.

KAPITEL 4

Umwandlungsoptimierung

Dieses Kapitel umfasst die folgenden Themen:

- Umwandlungsoptimierung, 25
- Aggregat-Umwandlungsoptimierung, 25
- Expressionsoptimierung, 26
- Java-Umwandlungsoptimierung, 28
- Joiner-Umwandlungsoptimierung, 31
- Lookup-Umwandlungsoptimierung, 32
- Sortierer-Umwandlungsoptimierung, 34
- SQL-Umwandlungsoptimierung, 35
- Umwandlungs-Cache, 37
- Eliminierung von Umwandlungsfehlern, 37
- Nebenwirkungen von Umwandlungen, 38
- Optimierung der Webdienstverbraucher-Umwandlung, 39

Umwandlungsoptimierung

Optimieren Sie Umwandlungen, damit der Data Integration Service Daten Umwandlungen in einem Mapping effizient verarbeiten kann.

Verwenden Sie zum Optimieren der Umwandlung die folgenden Optimierungsverfahren:

- Konfigurieren Sie Umwandlungen für die Optimierung.
- Eliminieren Sie Umwandlungsfehler.
- Konfigurieren Sie den Umwandlungs-Cache.

Aggregat-Umwandlungsoptimierung

Aggregat-Umwandlungen führen häufig zu einem Leistungsverlust, da Daten vor der Verarbeitung gruppiert werden müssen. Aggregat-Umwandlungen erfordern zusätzlichen Speicher für die Gruppenzwischenergebnisse.

Zum Beheben von Engpässen bei Aggregat-Umwandlungen können Sie die folgenden Lösungen verwenden:

Gruppieren Sie nach einfachen Spalten.

Durch das Gruppieren nach einfachen Spalten können Sie Aggregat-Umwandlungen optimieren. Verwenden Sie nach Möglichkeit in den Spalten für GROUP BY Zahlen anstelle von String- und Date-Werten. Vermeiden Sie in den Aggregat-Expressionen komplexe Expressionen.

Verwenden Sie sortierte Eingaben.

Sortieren Sie Daten für die Aggregat-Umwandlung, um die Mapping-Leistung zu erhöhen. Sortieren Sie Daten mit der Option „Sortierte Eingabe“.

Mit der Option „Sortierte Eingabe“ wird die Verwendung von Aggregat-Caches verringert. Wenn Sie die Option „Sortierte Eingabe“ verwenden, geht der Datenintegrationsdienst davon aus, dass sämtliche Daten nach Gruppen sortiert sind. Wenn der Datenintegrationsdienst Daten für eine Gruppe liest, werden Aggregat-Berechnungen ausgeführt. Ggf. werden Gruppeninformationen im Arbeitsspeicher gespeichert.

Mit der Option „Sortierte Eingabe“ reduzieren Sie die Menge der während des Mappings zwischengespeicherten Daten und verbessern die Leistung. Verwenden Sie die Option „Sortierte Eingabe“ oder eine Sortierer-Umwandlung, um sortierte Daten an die Aggregat-Umwandlung zu übergeben.

Wenn Sie die Option „Sortierte Eingabe“ in Mappings mit mehreren Partitionen verwenden, können Sie die Leistung erhöhen.

Filtern Sie Daten, bevor Sie sie aggregieren.

Wenn Sie eine Filter-Umwandlung im Mapping verwenden, platzieren Sie die Umwandlung vor die Aggregat-Umwandlung, um unnötige Aggregation zu vermeiden.

Beschränken Sie Port-Verbindungen.

Beschränken Sie die Anzahl der verbundenen Eingabe-/Ausgabe- oder Ausgabeports, um die Menge der Daten zu reduzieren, die die Aggregat-Umwandlung im Daten-Cache speichert.

Expressionsoptimierung

Einige verwendete Ausdrücke in einer Umwandlung können die Leistung reduzieren.

Zum Beheben von Expressionsengpässen können Sie die folgenden Lösungen verwenden:

Isolieren Sie langsame Expressionen.

Langsame Expressionen verringern die Mapping-Leistung. Um langsame Ausdrücke zu isolieren, entfernen Sie nacheinander die Ausdrücke aus dem Mapping und führen Sie das Mapping aus, um die Zeit zu bestimmen, die ohne die Ausdrücke benötigt wird. Wenn Sie einen erheblichen Unterschied bei der Mapping-Laufzeit feststellen, versuchen Sie, die langsame Expression zu optimieren.

Führen Sie die folgenden Schritte aus, um die Expressionsleistung zu bewerten:

1. Messen Sie die Zeit zum Ausführen des Mappings mit den ursprünglichen Expressionen.
2. Kopieren Sie das Mapping und ersetzen Sie die Hälfte der komplexen Expressionen durch eine Konstante.
3. Führen Sie das bearbeitete Mapping aus und messen Sie die Ausführungsdauer.
4. Erstellen Sie eine weitere Kopie des Mappings und ersetzen Sie die andere Hälfte der komplexen Expressionen durch eine Konstante.
5. Führen Sie das bearbeitete Mapping aus und messen Sie die Ausführungsdauer.

Klammern Sie allgemeine Logik aus.

Wenn das Mapping die gleiche Aufgabe an mehreren Stellen ausführt, verringern Sie die Anzahl der Ausführungen dieser Aufgabe, indem Sie die Aufgabe an einen früheren Zeitpunkt im Mapping verschieben. Angenommen, für ein Mapping sind fünf Target-Tabellen vorhanden. Für jedes Target muss ein Lookup nach Sozialversicherungsnummern ausgeführt werden. Statt das Lookup fünf Mal auszuführen, setzen Sie die Lookup-Umwandlung im Mapping vor die Stelle, an der sich der Datenfluss teilt. Übergeben Sie anschließend die Lookup-Ergebnisse an alle fünf Targets.

Minimieren Sie Aggregatfunktionsaufrufe.

Klammern Sie beim Schreiben von Expressionen so viele Aggregatfunktionsaufrufe wie möglich aus. Bei jedem Aufruf einer Aggregatfunktion muss der Data Integration Service Daten suchen und gruppieren. Beispielsweise liest der Data Integration Service in der folgenden Expression COLUMN_A, ermittelt die Summe, liest dann COLUMN_B, ermittelt die Summe und ermittelt schließlich die Summe der beiden Summen.

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

Wenn Sie die Aggregatfunktionen wie nachfolgend gezeigt ausklammern, addiert der Data Integration Service COLUMN_A zu COLUMN_B und ermittelt dann die Summe.

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

Ersetzen Sie häufige Expressionen durch lokale Variablen.

Wenn Sie in einer Umwandlung die gleiche Expression mehrmals verwenden, können Sie diese als lokale Variable festlegen. Sie können eine lokale Variable nur innerhalb der Umwandlung verwenden. Jedoch erhöhen Sie die Leistung, da die Variable nur einmal berechnet wird.

Wählen Sie numerische Operatoren statt String-Operatoren.

Der Data Integration Service verarbeitet numerische Operationen schneller als String-Operationen. Wenn Sie z. B. in den beiden Spalten EMPLOYEE_NAME und EMPLOYEE_ID große Mengen von Daten suchen und das Lookup für EMPLOYEE_ID konfigurieren, erhöhen Sie die Leistung.

Optimieren Sie CHAR-CHAR- und CHAR-VARCHAR-Vergleiche.

Wenn der Data Integration Service Vergleiche zwischen CHAR- und VARCHAR-Spalten ausführt, verringert sich die Geschwindigkeit jedes Mal, wenn in der Zeile abschließende Leerzeichen gefunden werden. Sie können die Option „TreatCHARasCHARonRead“ verwenden, wenn Sie den Data Integration Service im Informatica Administrator konfigurieren, damit der Data Integration Service abschließende Leerzeichen am Ende der Char-Quellfelder nicht abschneidet.

Wählen Sie DECODE statt LOOKUP.

Wenn Sie eine LOOKUP-Funktion verwenden, muss der Data Integration Service eine Tabelle in der Datenbank durchsuchen. Wenn Sie eine DECODE-Funktion verwenden, schließen Sie die Lookup-Werte in die Expression ein, so dass der Data Integration Service keine separate Tabelle durchsuchen muss. Verwenden Sie daher zum Erhöhen der Leistung DECODE, wenn Sie eine kleine Menge unveränderlicher Daten durchsuchen möchten.

Verwenden Sie Operatoren statt Funktionen.

Der Data Integration Service liest mit Operatoren geschriebene Expressionen schneller als Expressionen mit Funktionen. Verwenden Sie nach Möglichkeit zum Schreiben von Expressionen Operatoren. Beispielsweise enthält die folgende Expression geschachtelte CONCAT-Funktionen:

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

Sie können die Expression mit dem Operator || wie folgt umschreiben:

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

Optimieren Sie IIF-Funktionen:

IIF-Funktionen können einen Wert und eine Aktion zurückgeben. Dies ermöglicht kompaktere Expressionen. Angenommen, eine Quelle enthält drei Y/N-Flags: FLG_A, FLG_B, FLG_C. Werte sollen basierend auf den Werten der einzelnen Flags zurückgegeben werden.

Sie verwenden die folgende Expression:

```
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,
))))))))
```

Java-Umwandlungsoptimierung

Einige Java-Umwandlungen in einem Mapping können die Leistung reduzieren.

Berücksichtigen Sie die folgende Lösung, um die Java-Umwandlungsleistung zu erhöhen:
Aktivieren Sie die frühen Auswahl- oder push-into-Filteroptimierungsmethoden mit der Java-Umwandlung.

Sie können die frühe Auswahl- oder push-into-Optimierung in Java-Umwandlungen aktivieren. Aktualisieren Sie die Codeausschnitte auf der Registerkarte **Optimierungsschnittstellen** der Java-Umwandlung.

Frühe Auswahloptimierung mit der Java-Umwandlung

Sie können eine aktive oder passive Java-Umwandlung für die frühe Auswahloptimierung aktivieren, wenn die Java-Umwandlung keine Nebeneffekte aufweist. Der Optimierer übergibt der Filterlogik über die Java-Umwandlung und ändert die Filterbedingung nach Bedarf.

Zum Anzeigen der Codeausschnitte für die "Early Selection"-Optimierung wählen Sie PredicatePushOptimization im Navigator der Registerkarte **Optimierungsschnittstelle** aus.

allowPredicatePush

Boolescher Wert. Aktiviert "Early Selection". Ändern Sie die Funktion, um zur Aktivierung von "Early Selection" den Wahrheitswert "True" und eine Meldung zurückzugeben. Der Standardwert lautet "False" und die Funktion gibt eine Meldung mit dem Hinweis zurück, dass die Optimierung nicht unterstützt wird.

```
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

Boolescher Wert. Gibt an, ob die Java-Umwandlung einen Ausgabefeldfehler zurückgeben kann, beispielsweise einen Division durch Null-Fehler. Ändern Sie die Funktion, um "False" zurückzugeben, wenn die Java-Umwandlung keine Ausgabefeldfehler erzeugt. Wenn die Java-Umwandlung Feldfehler erzeugen kann, kann "Early Selection" vom Data Integration Service nicht verwendet werden.

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

getInputExpr

Gibt einen Informatica-Ausdruck zurück, der die Eingabewerte aus Eingabefeldern beschreibt, die ein Ausgabefeld enthalten. Der Optimierer muss wissen, welche Eingabefelder ein Ausgabefeld enthalten, um die Filterlogik durch die Umwandlung zu leiten.

```
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;
}
```

Ein Mapping enthält beispielsweise einen Filterausdruck, "`out0 > 8 `". Out0 ist der Wert des out0-Ausgabeports in der Java-Umwandlung. Sie können den Wert out0 als den Wert des in0-Eingabeports + 5 definieren. Der Optimierer kann den folgenden Ausdruck "`(in0 + 5) > 8`" mit der "Early Selection"-Optimierung hinter die Java-Umwandlung verschieben. Sie können NULL zurückgeben, wenn ein Ausgabefeld nicht über einen Eingabefeldausdruck verfügt. Der Optimierer verschiebt Filterausdrücke nicht hinter Ausgabefelder ohne Eingabeausdruck.
Sie können gegebenenfalls folgenden Code aufnehmen:

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

inputGroupsPushPredicateTo

Gibt eine Liste von Gruppen zurück, die die Filterlogik empfangen können. Die Java-Umwandlung verfügt über eine Eingabegruppe. Ändern Sie diese Funktion für die Java-Umwandlung nicht.

```
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;
    }
```

Push-Into-Optimierung mit der Java-Umwandlung

Sie können eine aktive Java-Umwandlung für die push-into-Optimierung aktivieren, wenn sie keine Nebeneffekte aufweist und die Optimierung keine Auswirkungen auf die Mapping-Ergebnisse hat.

Wenn Sie die push-into-Optimierung für die Java-Umwandlung konfigurieren, definieren Sie für die Java-Umwandlung eine Möglichkeit zum Speichern der Filterbedingung, die sie aus dem Optimierer enthält. Fügen Sie Code hinzu, der die Filterbedingung überprüft. Wenn die Java-Umwandlung die Filterlogik absorbieren kann, gibt die Java-Umwandlung eine wahre Bedingung an den Optimierer zurück. Der Optimierer entfernt die Filterumwandlung aus dem optimierten Mapping.

Beim Konfigurieren der Java-Umwandlung schreiben Sie den Code, der die Filterbedingung während der Optimierung als Umwandlungsmetadaten speichert. Sie schreiben den Code auch zum Abrufen der Filterbedingung zur Laufzeit sowie zum Entfernen der Zeilen entsprechend der Filterlogik.

Beim Definieren der Java-Umwandlung fügen Sie Code für die Push-Into-Optimierung auf der Registerkarte **Optimierungsschnittstelle** der Java-Umwandlung hinzu. Wählen Sie für den Zugriff auf Codeausschnitte für die Push-Into-Optimierung "FilterPushdownOptimization" im Navigator auf der Registerkarte **Optimierungsschnittstelle** der Umwandlung aus.

Das Developer Tool zeigt Codeausschnitte an, um Push-Into-Optimierung zu aktivieren und die Filterbedingung aus dem Optimierer abzurufen. Aktualisieren Sie die Codeausschnitte zum Aktivieren der Optimierung sowie zum Speichern der Filterlogik als Umwandlungsmetadaten.

isFilterSupported

Gibt TRUE zurück, um Push-Into-Optimierung zu aktivieren. Gibt FALSE zurück, um Push-Into-Optimierung zu deaktivieren.
Ändern Sie die Funktion zur Rückgabe von TRUE, um Push-Into-Optimierung zu aktivieren.

```
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

Erhält die Filterbedingung vom Optimierer.
Fügen Sie Code hinzu, um den Filter zu überprüfen und anzugeben, ob die Filterlogik in der Umwandlung verwendet werden kann. Wenn die Umwandlung den Filter absorbieren kann, verwenden Sie die folgende Methode zum Speichern der Filterbedingung als Umwandlungsmetadaten:

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

Der Schlüssel ist ein Bezeichner für die Metadaten. Sie können jeden String als Schlüssel definieren. Die Daten sind die Daten, die Sie speichern möchten, um festzustellen, welche Zeilen zur Laufzeit entfernt werden sollen. Beispielsweise kann es sich bei den Daten um die Filterbedingung handeln, die die Java-Umwandlung vom Optimierer erhält.

```
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");
}
```

Joiner-Umwandlungsoptimierung

Joiner-Umwandlungen können einen Leistungsabfall verursachen, da sie zur Laufzeit zusätzlichen Speicher für Zwischenergebnisse benötigen.

Zum Beheben von Engpässen bei Joiner-Umwandlungen können Sie die folgenden Lösungen verwenden:

Legen Sie die Master-Quelle als Quelle mit weniger doppelten Schlüsselwerten fest.

Wenn der Datenintegrationsdienst eine sortierte Joiner-Umwandlung verarbeitet, werden Zeilen für 100 eindeutige Schlüssel gleichzeitig zwischengespeichert. Wenn die Master-Quelle viele Zeilen mit demselben Schlüsselwert enthält, muss der Datenintegrationsdienst mehr Zeilen zwischenspeichern, was zu einer Verringerung der Leistung führen kann.

Festlegen der Master-Quelle als Quelle mit weniger Zeilen.

Die Joiner-Umwandlung vergleicht jede Zeile der Detail-Quelle mit der Master-Quelle. Je weniger Zeilen im Master enthalten sind, desto weniger Iterationen des Join-Vergleichs treten auf, was zu einer Beschleunigung des Join-Prozesses führt.

Durchführen von Joins in einer Datenbank, wenn möglich.

Das Durchführen eines Joins in einer Datenbank geht schneller als das Durchführen eines Joins während eines Mapping-Laufs. Der von Ihnen verwendete Typ des Datenbank-Joins kann sich auf die Leistung auswirken. Normale Joins sind schneller als Outer-Joins und resultieren in weniger Zeilen. In bestimmten Fällen kann in der Datenbank kein Join durchgeführt werden, beispielsweise das Zusammenführen von Tabellen aus zwei verschiedenen Datenbanken oder flachen Dateisystemen.

Zusammenfügen sortierter Daten, wenn möglich.

Konfigurieren Sie die Joiner-Umwandlung für die Verwendung sortierter Eingaben. Der Datenintegrationsdienst erhöht die Leistung durch Minimierung der Festplatteneingabe und Festplattenausgabe. Die größte Leistungserhöhung tritt auf, wenn Sie mit umfangreichen Datensätzen arbeiten. Legen Sie bei einer unsortierten Joiner-Umwandlung für die Quelle weniger Zeilen als für die Master-Quelle fest.

Optimieren Sie die Join-Bedingung.

Der Datenintegrationsdienst versucht, die Größe der Datenmenge eines Join-Operanden zu verringern, indem er die Zeilen aus der kleineren Gruppe liest, die übereinstimmenden Zeilen in der größeren Gruppe sucht und dann die Join-Operation durchführt. Das Verringern der Größe der Datenmenge verbessert die Mapping-Leistung, da der Datenintegrationsdienst nicht mehr unnötige Zeilen aus der größeren Quellgruppe liest. Der Datenintegrationsdienst verschiebt die Join-Bedingung in die größere Quellgruppe und liest nur die Zeilen, die mit der kleineren Gruppe übereinstimmen.

Verwenden Sie die Semi-Join-Optimierungsmethode.

Verbessern Sie mit der Semi-Join Optimierungsmethode die Mapping-Leistung, wenn eine Eingabegruppe über mehr Zeilen als die andere Gruppe verfügt und wenn die größere Gruppe viele Zeilen enthält, für die in der kleineren Gruppe keine Übereinstimmung gemäß der Join-Bedingung vorhanden ist.

Lookup-Umwandlungsoptimierung

Lookup-Umwandlungen können je nach Cache-Typ und Lookup-Bedingungen die Leistung verringern.

Zum Beheben von Engpässen bei Lookup-Umwandlungen können Sie die folgenden Lösungen verwenden:

Verwenden Sie den optimalen Datenbanktreiber.

Der Data Integration Service kann mit einem nativen Datenbanktreiber oder einem ODBC-Treiber eine Verbindung mit einer Lookup-Tabelle herstellen. Native Datenbanktreiber bieten eine bessere Mapping-Leistung als ODBC-Treiber.

Speichern Sie Lookup-Tabellen für relationale oder Einfachdatei-Lookups zwischen.

Um die Lookup-Leistung für relationale oder Einfachdateiquellen zu verbessern, aktivieren Sie Lookup-Caching in der Umwandlung. Wenn Sie Caching aktivieren, speichert der Datenintegrationsdienst die Lookup-Tabelle. Wenn Sie das Mapping ausführen, fragt der Datenintegrationsdienst nicht die Lookup-Tabelle, sondern den Lookup-Cache ab. Wenn diese Option nicht aktiviert ist, fragt der Data Integration Service die Lookup-Tabelle Zeile für Zeile ab.

Das Ergebnis der Lookup-Abfrage und -Verarbeitung ist unabhängig davon, ob Sie die Lookup-Tabelle zwischenspeichern, dasselbe. Für kleinere Lookup-Tabellen kann jedoch die Mapping-Leistung durch Verwendung eines Lookup-Caches verbessert werden. Im Allgemeinen sollten Sie Lookup-Tabellen zwischenspeichern, die weniger als 300 MB erfordern.

Speichern Sie Lookup-Tabellen für Lookups logischer Datenobjekte zwischen.

Um die Lookup-Leistung für ein logisches Datenobjekt zu verbessern, können Sie Datenobjekt-Caching für den Datenintegrationsdienst aktivieren. Wenn Sie Datenobjekt-Caching aktivieren, speichert der Datenintegrationsdienst das logische Datenobjekt. Um Datenobjekt-Caching zu aktivieren, müssen Sie das Mapping für eine Anwendung bereitstellen, Caching des logischen Datenobjekts aktivieren und das Mapping mit dem Befehl `infacmd ms runmapping` ausführen. Wenn Sie das Mapping ausführen, fragt der Datenintegrationsdienst nicht das logische Datenobjekt, sondern den Datenobjekt-Cache ab.

Wenn Sie das Mapping über das Developer-Tool ausführen, fragt die Lookup-Umwandlung das logische Datenobjekt Zeile für Zeile ab.

Verwenden Sie den entsprechenden Cache-Typ.

Verwenden Sie zum Erhöhen der Leistung die folgenden Typen von Caches:

- Gemeinsam genutzter Cache. Sie können den Lookup-Cache für mehrere Umwandlungen gemeinsam nutzen. Sie können einen unbenannten Cache gemeinsam für Umwandlungen in demselben Mapping nutzen. Sie können einen benannten Cache gemeinsam für Umwandlungen in demselben Mapping oder in unterschiedlichen Mappings nutzen.
- Persistenter Cache. Um die Cache-Dateien zu speichern und wiederzuverwenden, können Sie die Umwandlung zur Verwendung eines persistenten Cache konfigurieren. Verwenden Sie diese Funktion, wenn Sie wissen, dass die Lookup-Tabelle zwischen Mapping-Ausführungen unverändert bleibt. Die Verwendung eines persistenten Caches kann die Leistung verbessern, da der Data Integration Service den Arbeitsspeicher-Cache aus den Cache-Dateien und nicht aus der Datenbank erstellt.

Aktivieren Sie gleichzeitige Caches.

Wenn der Data Integration Service Mappings verarbeitet, die Lookup-Umwandlungen enthalten, und die erste Datenzeile in einer zwischengespeicherten Lookup-Umwandlung verarbeitet, erstellt er einen Cache im Arbeitsspeicher. Wenn ein Mapping mehrere Lookup-Umwandlungen enthält und die erste Datenzeile von der Lookup-Umwandlung verarbeitet wurde, erstellt der Data Integration Service die Caches nacheinander. Dies verlangsamt die Verarbeitung der Lookup-Umwandlungen.

Zum Verbessern der Leistung können Sie gleichzeitige Caches aktivieren. Wenn die Anzahl der zusätzlichen gleichzeitigen Pipelines auf 1 oder einen höheren Wert festgelegt ist, erstellt der Data Integration Service Caches nicht nacheinander, sondern gleichzeitig. Die Leistung wird erheblich verbessert, wenn die Mappings mehrere aktive Umwandlungen enthalten, deren Ausführung möglicherweise lange dauert, wie dies z. B. bei Aggregat-, Joiner- oder Sorter-Umwandlungen der Fall ist. Wenn Sie mehrere gleichzeitige Pipelines aktivieren, wartet der Data Integration Service nicht mehr auf den Abschluss aktiver Mappings, bevor er den Cache erstellt. Andere Lookup-Umwandlungen in der Pipeline erstellen ebenfalls gleichzeitig Caches.

Optimieren Sie Lookup-Bedingungs-Matches.

Wenn die Lookup-Umwandlung Lookup-Cache-Daten mit der Lookup-Bedingung abgleicht, sortiert und ordnet sie die Daten, um den ersten und den letzten übereinstimmenden Wert zu bestimmen. Sie können die Umwandlung für die Rückgabe eines beliebigen Wertes, der mit der Lookup-Bedingung übereinstimmt, konfigurieren. Wenn Sie die die Umwandlung für die Rückgabe eines beliebigen übereinstimmenden Wertes konfigurieren, gibt die Umwandlung den ersten Wert zurück, der mit der Lookup-Bedingung übereinstimmt. Sie indiziert nicht alle Ports, wie dies der Fall ist, wenn Sie die Umwandlung für die Rückgabe des ersten oder letzten übereinstimmenden Wertes konfigurieren.

Wenn Sie einen beliebigen übereinstimmenden Wert verwenden, kann dies die Leistung verbessern, da die Umwandlung nicht alle Ports indiziert, wodurch die Leistung beeinträchtigt werden kann.

Verringern Sie die Anzahl zwischengespeicherter Zeilen.

Sie können die Anzahl der Zeilen im Cache verringern, um die Leistung zu erhöhen. Verwenden Sie die Option „Lookup-SQL-Überschreibung“, um der Standard-SQL-Anweisung eine WHERE-Klausel hinzuzufügen. Wenn Sie eine WHERE-Klausel zur Lookup-Umwandlung hinzufügen, die einen dynamischen Cache verwendet, verwenden Sie vor der Lookup-Umwandlung eine Filterumwandlung, um Zeilen in den dynamischen Cache zu übergeben, die mit der WHERE-Klausel übereinstimmen.

Überschreiben Sie die ORDER BY-Anweisung.

Der Data Integration Service generiert standardmäßig eine ORDER BY-Anweisung für ein zwischengespeichertes Lookup. Die ORDER BY-Anweisung enthält alle Lookup-Ports. Um die Leistung zu erhöhen, unterdrücken Sie die ORDER BY-Anweisung und geben Sie eine Überschreibung der ORDER BY-Anweisung mit weniger Zeilen ein.

Der Data Integration Service generiert immer eine ORDER BY-Anweisung, auch wenn Sie in der Überschreibung nur eine Zeile eingeben. Fügen Sie nach der ORDER BY-Überschreibung zwei Bindestriche („--“) ein, um die generierte ORDER BY-Anweisung zu unterdrücken.

Beispiel: In einer Lookup-Umwandlung wird die folgenden Lookup-Bedingung verwendet:

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

Die Lookup-Umwandlung enthält drei im Mapping verwendete Lookup-Ports: ITEM_ID, ITEM_NAME und PRICE Wenn Sie die ORDER BY-Anweisung eingeben, geben Sie die Spalten in derselben Reihenfolge wie die Ports in der Lookup-Bedingung ein. Sie müssen außerdem alle reservierten Datenbankwörter in Anführungszeichen einschließen.

Geben Sie in der Lookup-SQL-Überschreibung die folgende Lookup-Abfrage ein:

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

Verwenden Sie einen Computer mit größerem Arbeitsspeicher.

Um die Mapping-Leistung zu erhöhen, führen Sie das Mapping auf einem Data Integration Service-Knoten mit einem großen Umfang an Arbeitsspeicher aus. Legen Sie die Größe des Index- und Daten-Caches auf einen möglichst hohen Wert fest, ohne den Computer zu überfordern. Wenn der Data Integration Service-

Knoten über genügend Arbeitsspeicher verfügt, vergrößern Sie den Cache, damit alle Daten im Arbeitsspeicher gespeichert werden können, ohne Daten auf den Datenträger auszulagern.

Optimieren Sie die Lookup-Bedingung.

Wenn Sie mehrere Lookup-Bedingungen einschließen, platzieren Sie diese in der folgenden Reihenfolge, um die Lookup-Leistung zu optimieren:

- Gleich (=)
- Kleiner als (<), größer als (>), kleiner oder gleich (< =), größer als oder gleich (> =)
- Nicht gleich (!=)

Filtern Sie Lookup-Zeilen.

Erstellen Sie zum Verbessern der Leistung eine Filterbedingung, um die Anzahl der Lookup-Zeilen zu verringern, die beim Erstellen des Lookup-Caches aus der Quelle abgerufen werden.

Indizieren Sie die Lookup-Tabelle.

Der Data Integration Service muss Werte in den Lookup-Bedingungsspalten abfragen, sortieren und vergleichen. Der Index muss jede in einer Lookup-Bedingung verwendete Spalte enthalten.

Sie können die Leistung für folgende Typen von Lookups verbessern:

- Zwischengespeicherte Lookups. Indizieren Sie zum Verbessern der Leistung die Spalten in der ORDER BY-Anweisung des Lookups. Die Mapping-Log-Datei enthält die ORDER BY-Anweisung.
- Nicht zwischengespeicherte Lookups. Indizieren Sie zum Verbessern der Leistung die Spalten in der Lookup-Bedingung. Der Data Integration Service gibt für jede an die Lookup-Umwandlung übergebene Zeile eine SELECT-Anweisung aus.

Optimieren Sie mehrere Lookups.

Wenn ein Mapping mehrere Lookups enthält, können diese die Leistung beeinträchtigen, selbst wenn der Cache aktiviert und genügend Heap-Speicher vorhanden ist. Optimieren Sie zum Verbessern der Gesamtleistung die Lookup-Umwandlungen, die die größte Menge an Daten abfragen.

Wenn sich im Mapping die Lookup-Tabelle in derselben Datenbank wie die Quelltabelle befindet und Zwischenspeichern nicht möglich ist, verknüpfen Sie die Tabellen in der Quelldatenbank, statt eine Lookup-Umwandlung zu verwenden.

Sortierer-Umwandlungsoptimierung

Sortiervorgang-Umwandlungen können die Leistung verringern, wenn der physische RAM auf dem Datenintegrationsdienst-Knoten nicht über ausreichend Arbeitsspeicher verfügt, um einen Sortiervorgang auszuführen.

Zum Beheben von Engpässen bei der Sortierer-Umwandlung können Sie die folgende Lösung verwenden:

Reservieren Sie genügend Speicher.

Um eine optimale Leistung zu erzielen, legen Sie als Sortierer-Cachegröße einen Wert fest, der kleiner oder gleich dem verfügbaren physischen RAM auf dem Datenintegrationsdienst-Knoten ist. Reservieren Sie mindestens 16 MB physischen Speicher zum Sortieren von Daten mit der Sortierer-Umwandlung zu. Die Standardgröße des Sortierer-Caches ist 16.777.216 Bytes. Wenn der Datenintegrationsdienst nicht genügend Speicher zum Sortieren der Daten reservieren kann, schlägt das Mapping fehl.

Wenn die Menge der eingehenden Daten größer ist als die Menge der Sortierer-Cachegröße, speichert der Datenintegrationsdienst die Daten temporär im Arbeitsverzeichnis der Sortierer-Umwandlung. Beim

Speichern von Daten im Arbeitsverzeichnis benötigt der Datenintegrationsdienst einen Festplattenspeicher, der mindestens doppelt so groß ist, wie die eingehende Datenmenge.

SQL-Umwandlungsoptimierung

Immer wenn der Datenintegrationsdienst eine neue Abfrage verarbeitet, ruft er die Funktion SQLPrepare auf, um eine SQL-Prozedur zu erstellen und an die Datenbank zu übergeben. Wenn sich die Abfrage für jede Eingabezeile ändert, kann sich die Leistung reduzieren.

Zum Beheben von Engpässen bei Lookup-Umwandlungen können Sie die folgenden Lösungen verwenden:

Verwenden Sie in einer SQL-Umwandlungsabfrage keine Transaktionsanweisungen.

Wenn eine SQL-Abfrage Commit- und Rollback-Abfrageanweisungen enthält, muss der Datenintegrationsdienst die SQL-Prozedur nach jedem Commit bzw. Rollback neu erstellen. Verwenden Sie zum Optimieren der Leistung keine Transaktionsanweisungen in einer SQL-Umwandlungsabfrage.

Aktivieren Sie die frühe Auswahl oder die push-into-Filteroptimierungsmethoden mit der SQL-Umwandlung.

Um die Leistung zu erhöhen, können Sie die frühe Auswahl oder Push-into-Optimierungsmethode mit der SQL-Umwandlung aktivieren.

"Early Selection"-Optimierung mit der SQL-Umwandlung

Der Data Integration Service kann "Early Selection"-Optimierung mit einer SQL-Umwandlung durchführen, wenn die Filterbedingung ausschließlich auf Pass-Through-Ports verweist und die SQL-Umwandlung keine Nebeneffekte aufweist.

Die SQL-Umwandlung weist unter folgenden Umständen Nebeneffekte auf:

- Die SQL-Abfrage aktualisiert eine Datenbank. Die SQL-Abfrage enthält eine Anweisung, wie beispielsweise CREATE, DROP, INSERT, UPDATE, GRANT oder REVOKE.
- Die Umwandlung gibt Leerzeilen für SELECT-Anweisungen zurück, die keine Ergebnisse liefern. Die Zeilen können unter Umständen Pass-Through-Portwerte, SQL-Fehlerinformationen oder das NUMRowsAffected-Feld enthalten.

Aktivieren der frühen Auswahloptimierung mit der SQL-Umwandlung

Aktivieren Sie die frühe Auswahloptimierung in der SQL-Umwandlung, wenn die SQL-Umwandlung keine Nebeneffekte aufweist.

1. Aktivieren Sie die Option **Nur Datenbank-Ausgabe zurückgeben** unter **Erweiterte Eigenschaften** in der SQL-Umwandlung
2. Löschen Sie **Hat Nebeneffekte** in der Umwandlung **Erweiterte Eigenschaften**.
3. Weist die Umwandlung einen **NumAffectedRows**-Port auf, entfernen Sie den Port.

Push-Into-Optimierung mit der SQL-Umwandlung

Mit der Push-Into-Optimierung verschiebt der Data Integration Service die Filterlogik aus einer Filterumwandlung im Mapping in die Abfrage in der SQL-Umwandlung.

Verwenden Sie die folgenden Regeln und Richtlinien, wenn Sie die Push-Into-Optimierung mit der SQL Umwandlung aktivieren:

- Die SQL-Umwandlungsabfrage darf nur „AUSWAHL“-Anweisungen beinhalten.
- Die SQL-Umwandlungsabfrage muss eine gültige Unterabfrage sein.
- Die Filterbedingung kann nicht auf die Felder "SQL-Fehler" oder "NumRowsAffected" verweisen.
- Die Namen der Ausgabeports müssen mit den Namen der Spalten in der SQL-SELECT-Anweisung übereinstimmen. Wenn Sie auf einen Ausgabeport in einer Filterbedingung verweisen, verschiebt der Data Integration Service den entsprechenden Portnamen in die SQL-Abfrage. Sie können der SQL Aliase hinzufügen, wenn die Spalten in der Abfrage nicht mit den Namen der Ausgabeports übereinstimmen. Beispiel: `SELECT mycolname1 AS portname1, mycolname2 AS portname2`.
- Die Umwandlung darf keine Nebeneffekte aufweisen.

Beispiel für eine Push-Into-Optimierung mit der SQL-Umwandlung

Eine SQL-Umwandlung ruft Aufträge nach Kunden-ID ab. Eine Filterumwandlung, die nach der SQL-Umwandlung erscheint, gibt nur die Zeilen zurück, bei denen die Bestellmenge mehr als 1000 beträgt.

Der Data Integration Service fügt den folgenden Filter in eine SELECT-Anweisung in der SQL-Umwandlung ein:

```
orderAmount > 1000
```

Jede Anweisung in der SQL-Abfrage wird zu einer separaten Unterabfrage der SELECT-Anweisung, die den Filter enthält.

Die folgende Abfrageanweisung zeigt die ursprüngliche Abfrageanweisung als Unterabfrage in der SELECT-Anweisung an:

```
SELECT <customerID>, <orderAmount>, … FROM (original query statements) ALIAS WHERE
<orderAmount> > 1000
```

Wenn die SQL-Abfrage mehrere Anweisungen enthält, wird jede Anweisung in eine separate Unterabfrage eingeschlossen. Der Unterabfrage hat dieselbe Syntax, einschließlich der WHERE-Klausel.

Die Ports *customerID* und *orderAmount* sind die Namen der Ausgabeports in der SQL-Umwandlung. Die Unterabfrage enthält weder Pass-Through-Ports noch den SQL-Fehler noch die statistischen SQL-Ports. Wenn Sie mehrere Filter in die SQL-Umwandlung einfügen, enthält die WHERE-Klausel alle Filter.

Aktivieren der Push-Into-Optimierung mit der SQL-Umwandlung

Aktivieren Sie die Push-Into-Optimierung, indem Sie Eigenschaften auf der Registerkarte **Erweiterte Eigenschaften** der SQL-Umwandlung konfigurieren.

1. Löschen Sie **Hat Nebeneffekte**.
2. Aktivieren Sie **Nur Datenbank-Ausgabe zurückgeben**.
3. Setzen Sie·**Max. Ausgabe des Zeilenzählers** auf Null.
4. Aktivieren Sie die Push-Into-Optimierung.

Umwandlungs-Cache

Wenn Sie ein Mapping ausführen, das eine Aggregator-, Joiner- Lookup-, Rang- oder Sortiererumwandlung verwendet, erstellt der Datenintegrationsdienst Caches im Arbeitsspeicher, um die Umwandlung zu verarbeiten. Wenn der Datenintegrationsdienst mehr Speicherplatz erfordert, speichert er überzählige Werte in Cache-Dateien auf dem Datenträger.

Zum Beheben von Engpässen beim Umwandlungs-Cache können Sie die folgende Lösung verwenden:

Legen Sie fest, dass die Umwandlungen genügend Speicherplatz reservieren, um den Cache im Arbeitsspeicher zu speichern.

Um die Verarbeitungszeit für die Aggregat-, Joiner-, Lookup-, Rang- oder Sortiererumwandlung zu verringern, legen Sie fest, dass die Umwandlungen ausreichend Speicherplatz zum Speichern des Cache im Arbeitsspeicher reservieren. Wenn Sie einen Umfang an Cache-Arbeitsspeicher festlegen, der größer oder gleich dem Umfang ist, der zum Zwischenspeichern der Daten und des Index benötigt wird, erhöhen Sie die Leistung, indem Sie den E/A-Aufwand des Systems verringern. Wenn der Datenintegrationsdienst Cache-Dateien auf den Datenträger schreibt, erhöht sich aufgrund des E/A-Aufwands des Systems die Verarbeitungszeit.

Standardmäßig konfiguriert der Datenintegrationsdienst die Cache-Arbeitsspeicheranforderungen automatisch zur Laufzeit. Nach dem Ausführen eines Mappings im automatischen Cache-Modus können Sie die Cachegrößen für die Umwandlungen optimieren. Sie analysieren die Umwandlungsstatistiken im Mapping-Protokoll, um die für die Verarbeitung der Umwandlungen erforderlichen Cachegrößen im Arbeitsspeicher zu bestimmen. Wenn Sie die Cachegröße zur Verwendung des im Mapping-Protokoll angegebenen Werts konfigurieren, können Sie sicherstellen, dass kein zugeordneter Arbeitsspeicher ungenutzt bleibt. Die optimale Cachegröße variiert jedoch basierend auf der Größe der Quelldaten. Überprüfen Sie die Mapping-Protokolle nach der Ausführung weiterer Mappings, um Änderungen an der Cachegröße zu verfolgen. Wenn Sie eine bestimmte Cachegröße für eine wiederverwendbare Umwandlung konfigurieren, stellen Sie sicher, dass die Cachegröße optimal für jede Verwendung der Umwandlung in einem Mapping ist.

Eliminierung von Umwandlungsfehlern

Bei großen Zahlen verringern Umwandlungsfehler die Leistung des Data Integration Service. Der Data Integration Service wird bei jedem Umwandlungsfehler angehalten, um die Fehlerursache zu bestimmen und die Zeile aus dem Datenfluss zu entfernen, die den Fehler verursacht. Anschließend schreibt der Data Integration Service normalerweise die Zeile in die Mapping-Protokolldatei der Dienstprotokolle für die Datenintegration.

Zum Beheben von Engpässen aufgrund von Umwandlungsfehlern können Sie die folgenden Lösungen verwenden:

Überprüfen Sie in der Mapping-Log-Datei, wo die Umwandlungsfehler auftreten, und untersuchen Sie die entsprechenden Umwandlungsbeschränkungen.

Umwandlungsfehler treten auf, wenn im Data Integration Service Konvertierungsfehler, Konflikte erzeugende Mapping-Logik oder eine als Fehler konfigurierte Bedingung, z. B. keine Eingabe, auftreten. Überprüfen Sie in der Mapping-Log-Datei, wo die Umwandlungsfehler auftreten. Wenn die Fehler in bestimmten Umwandlungen gehäuft auftreten, untersuchen Sie die entsprechenden Umwandlungsbeschränkungen.

Konfigurieren Sie eine niedrigere Tracingebene.

Wenn Sie ein Mapping ausführen müssen, das eine große Anzahl von Umwandlungsfehlern erzeugt, lässt sich die Leistung durch Festlegen einer niedrigeren Tracingebene verbessern. Dies wird jedoch nicht als langfristige Lösung für Umwandlungsfehler empfohlen.

Nebenwirkungen von Umwandlungen

Eine Umwandlung weist Nebeneffekte auf, wenn sie Zeilen zurückgibt und ein Objekt ändert, oder wenn sie mit anderen Objekten oder Funktionen interagiert. Die Umwandlung kann eine Datenbank ändern, einen Betrag zu einer Summe addieren, eine Ausnahme auslösen, eine E-Mail schreiben oder andere Funktionen mit Nebenwirkungen aufrufen.

Der Datenintegrationsdienst bestimmt, welche Umwandlungen Nebenwirkungen haben, bevor er ein Mapping optimiert. Der Datenintegrationsdienst geht davon aus, dass eine Umwandlung Nebenwirkungen hat, wenn er nicht bestimmen kann, ob eine Umwandlung Nebenwirkungen hat.

Umwandlungen mit Nebenwirkungen beschränken die Möglichkeiten des Datenintegrationsdienstes zum Optimieren des Mappings. Die frühe Auswahloptimierung, die Verzweigungsbereinigungsoptimierung, die globale Vorhersageoptimierung und die Push-Into-Optimierung verändern die Mapping-Ergebnisse, wenn sie vom Datenintegrationsdienst auf eine Umwandlung mit Nebeneffekten angewendet werden. Bei der frühen Auswahloptimierung und der Push-Into-Optimierung wird Filterlogik aus einer Filter-Umwandlung so nahe wie möglich an die Quelle verschoben. Wenn der Filter vor der Nebenwirkung angewendet wird, ändern sich die Mapping-Ergebnisse.

Beispiel: Eine Umwandlung empfängt eine Kunden-ID und gibt Zeilen mit Bestellungsinformationen zurück. Die Umwandlung schreibt außerdem die Bestellungen in eine Datei. Wenn der Datenintegrationsdienst eine Filteroptimierung anwendet, bevor er Bestellungen in die Datei schreibt, empfängt die Datei weniger Zeilen als bei einer Filterung zu einem späteren Zeitpunkt im Mapping. Die Nebenwirkung der Umwandlung ist die Funktion zum Schreiben von Bestellungsdatensätzen in eine Datei.

Die folgenden Umwandlungen haben Nebenwirkungen:

- SQL-Umwandlung, Webdienstverbraucher-Umwandlung und Java-Umwandlung, sofern die Nebenwirkungseigenschaft nicht deaktiviert ist.
- Umwandlungen, die eine ABORT()- oder ERROR()-Funktion aufrufen, E-Mails senden oder eine gespeicherte Prozedur aufrufen können.
- Umwandlungen, die in Dateien oder Datenbanken schreiben.
- Umwandlungen, die eine Anzahl in einem Variablen-Port speichern Beispiel: COUNT=COUNT+1.

Die SQL-Umwandlung, die Webdienstverbraucher-Umwandlung und die Java-Umwandlung haben standardmäßig Nebenwirkungen. Wenn Sie festlegen, dass die Umwandlung Zeilen ohne Nebenwirkungen verarbeiten soll, können Sie in **Erweiterte Eigenschaften** die Eigenschaft **Hat Nebenwirkungen** deaktivieren. Wenn die Umwandlung keine Nebenwirkungen hat, können Sie Optimierung aktivieren, indem Sie in diesen Umwandlungen zusätzliche Eigenschaften konfigurieren.

Optimierung der Webdienstverbraucher-Umwandlung

Die Webdienstkunden-Umwandlung kann die Leistung verringern, wenn ein Mapping den Webdienst mehrfach aufruft.

Berücksichtigen Sie die folgenden Lösungen zu Engpässen bei Webdienstkunden-Umwandlungen:

Konfigurieren Sie die Webdienstkunden-Umwandlung, um die Cookie-Authentifizierung zu verwenden.

Der Server des Remote-Webdienstes verfolgt die Webdienstkunden-Nutzer anhand der Cookies. Sie können die Leistung erhöhen, wenn ein Mapping einen Webdienst mehrmals aufruft.

Wenn Sie den Cookie-Port für eine Webdienst-Anfragenachricht projektieren, gibt der Webdienstprovider einen Cookie-Wert in der Antwortnachricht zurück. Sie können den Cookie-Wert an eine im Mapping nachgelagerte Umwandlung übergeben oder den Cookie-Wert in einer Datei speichern. Beim Speichern des Cookie-Werts in einer Datei können Sie den Cookie als Eingabe in die Webdienstkunden-Umwandlung konfigurieren. Sie können den Cookie-Ausgabeport für alle Ausgabegruppen der Webdienstkunden-Umwandlung projektieren.

Aktivieren Sie die frühe Auswahl oder die push-into-Filteroptimierungsmethoden mit der Webdienstkunden-Umwandlung.

Um die Leistung zu erhöhen, kann der Datenintegrationsdienst die frühe Auswahl oder die Push-into-Optimierungsmethode mit der Webdienstverbraucher-Umwandlung anwenden. Zum Anwenden der frühen Auswahloptimierung, darf der Webdienst keine Nebeneffekte aufweisen und darf Unrichtigkeiten nicht als Fehler behandeln. Um die push-into-Optimierung anzuwenden, darf der Webdienst keine Nebeneffekte aufweisen, darf Unrichtigkeiten nicht als Fehler behandeln und die Filterbedingung muss auf pass-through-Ports zugreifen.

Der Webdienst hat Nebeneffekte, wenn er außer dem Zurückgeben einer Antwort an die Webdienstkunden-Umwandlung noch andere Funktionen ausführt. Der Webdienst hat Nebeneffekte, wenn er eine Datenbank ändert, in eine Datei schreibt, E-Mails schreibt, einen Zähler aktualisiert oder andere Webdienste mit Nebeneffekten aufruft.

Frühe Auswahloptimierung mit der Webdienstkunden-Umwandlung

Wenn der Data Integration Service die frühe Auswahloptimierungs-Methode mit der Webdienstkunden-Umwandlung anwendet, verschiebt dieser die Filterbedingungen vor die Webdienstkunden-Umwandlung im Mapping und damit näher an die Quelle.

Aktivieren der "Early Selection"-Optimierung mit der Webdienst-Verbraucher-Umwandlung

Aktivieren Sie die "Early Selection"-Optimierung für die Webdienst-Verbraucher-Umwandlung, wenn die Umwandlung keine Nebeneffekte hat und keine "Faults" als Fehler behandelt.

1. Öffnen Sie die Ansicht **Erweiterte Eigenschaften** der Web-Dienst-Verbraucher-Umwandlung.
2. Deaktivieren Sie **"Fault" als Fehler behandeln**.
3. Deaktivieren Sie **Hat Nebeneffekte**.

Push-Into-Optimierung mit der Web-Dienst-Verbraucher-Umwandlung

Sie können die Push-into-Optimierung mit der Webdienstverbraucherumwandlung konfigurieren, wenn die Umwandlung sich in einer virtuellen Tabelle in einem SQL-Datendienst befindet.

Das Mapping ruft den Web-Dienst auf, um einen Datensatz oder eine Teilmenge der Daten basierend auf den Anweisungen in der SQL-Abfrage des Endbenutzers abzurufen. Die SQL-Abfrage des Endbenutzers enthält eine optionale Filterbedingung.

Mit der Push-Into-Optimierung erhält die Web-Dienst-Verbraucher-Umwandlung den Filterwert in einem Filterport. Der Filterport ist ein nicht verbundener Eingabeport, den Sie als Filterport angeben, wenn Sie die Push-Into-Optimierung konfigurieren. Der Filterport verfügt über einen Standardwert, der sicherstellt, dass der Web-Dienst alle Zeilen zurückgibt, wenn die Endbenutzerabfrage keine Filter enthält. Der Filterport ist kein Übergabeport.

Hinweis: Das Filterfeld muss Teil der Root-Gruppe in der Web-Dienst-Anfrage sein.

Wenn Sie einen Filterport konfigurieren, geben Sie einen Ausgabeport in der Web-Dienst-Verbraucher-Umwandlung an, der die Spaltendaten aus der Web-Dienst-Antwort empfängt. Ist der Filterport beispielsweise ein Eingabeport mit der Bezeichnung EmployeeID, handelt es sich bei dem Ausgabeport aus der Antwort unter Umständen um einen Port mit der Bezeichnung EmployeeNum. Das Developer Tool muss den Eingabefilterport und einen Ausgabeport zuordnen, um die Filterlogik aus den virtuell gelesenen Daten in die Web-Dienst-Verbraucher-Anfrage zu verschieben. Die Eingabeports für eine Web-Dienst-Anfrage unterscheiden sich in der Regel von den Ausgabeports aus der Web-Dienst-Antwort.

Der Filterfeld kann kein Übergabeport sein. Wenn Sie einen Filterport konfigurieren, ändert sich der Standardwert des Ports in den Wert der Filterbedingung, wodurch der Wert des Pass-Through-Ausgabeports geändert wird. Ein Filter, der auf dem Pass-Through-Ausgabeport basiert, gibt unerwartete Ergebnisse zurück.

Sie können mehrere Filterausdrücke in die Web-Dienst-Verbraucher-Umwandlung verschieben. Jede Filterbedingung muss das folgende Format aufweisen:

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

Die Filterbedingungen müssen durch AND verbunden werden. Die Bedingungen können nicht durch OR verbunden werden.

Beispiel für eine Push-Into-Optimierung mit der Web-Dienst-Verbraucher-Umwandlung

Ein SQL-Datendienst gibt Aufträge für alle Kunden oder Aufträge für einen bestimmten Kunden basierend auf der SQL-Abfrage zurück, die er vom Benutzer erhält.

Der Datendienst enthält ein logisches Datenobjekt mit den folgenden Komponenten:

Kundentabelle

Eine Oracle-Datenbanktabelle, die Kundeninformationen enthält.

Web-Dienst-Verbraucher-Umwandlung

Eine Umwandlung, die einen Web-Dienst aufruft, um die aktuellen Kundenaufträge abzurufen. Die Web-Dienst-Verbraucher-Umwandlung verfügt über Ports für die Kunden-ID (customerID) oder die Auftragsnummer (orderNum). Die Umwandlung verfügt über Übergabeports mit Kundendaten, die aus der Kundentabelle stammen. Der Port für die Auftragsnummer (orderNum) fungiert als Filterport und ist nicht verbunden. orderNum weist den Standardwert "*" auf. Wenn der Web-Dienst diesen Wert in der Web-Dienst-Anfrage empfängt, werden alle Aufträge zurückgegeben.

Virtuelle Tabelle für Aufträge

Hierbei handelt es sich um eine virtuelle Tabelle, die Kunden- und Auftragsdaten vom Webdienst empfängt. Der Endbenutzer verwendet diese Tabelle für Anfragen. Die Tabelle für Aufträge enthält eine Spalte für Kunden, eine Spalte für die Kunden-ID sowie Kunden- und Auftragsdaten.

Der Endbenutzer übergibt die folgenden SQL-Abfrage an den SQL-Datendienst:

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

Der Data Integration Service teilt die Abfrage auf, um das Mapping zu optimieren. Der Data Integration Service verwendet "Early Selection"-Optimierung und verschiebt die Filterlogik `customer = 23` in die gelesenen Kundendaten. Der Data Integration Service verwendet Push-Into-Optimierung und verschiebt die Filterlogik `orderID = 56` in den Filterport der Web-Dienst-Verbraucher-Umwandlung. Die Web-Dienst-Verbraucher-Umwandlung ruft die Auftrags-ID 56 für den Kunden 23 ab.

Aktivieren der Push-Into-Optimierung mit der Web-Dienst-Verbraucher-Umwandlung

Aktivieren Sie die Push-Into-Optimierung für die Webdienst-Verbraucher-Umwandlung, wenn die Umwandlung keine Nebeneffekte hat und keine "Faults" als Fehler behandelt.

1. Öffnen Sie die Ansicht **Erweiterte Eigenschaften** der Web-Dienst-Verbraucher-Umwandlung.
2. Deaktivieren Sie **"Fault" als Fehler behandeln**.
3. Deaktivieren Sie **Hat Nebeneffekte**.
4. Klicken Sie auf die Schaltfläche **Öffnen** in der Eigenschaft **Push-Into-Optimierung**.
5. Wählen Sie den Namen des Filterports im Dialogfeld "Optimierte Eingabe" aus.
   Sie können mehrere Filterports auswählen.
6. Klicken Sie auf die Spalte **Ausgabe**.
7. Wählen Sie für jeden Filterport den Ausgabeport aus, der die gefilterte Spalte in der Web-Dienst-Antwort enthält.
8. Geben Sie einen Standardwert für jeden Filterport ein.

   Hinweis: Sie können einen Standardwert für einen Web-Dienst-Verbraucher-Port konfigurieren, es sei denn, es handelt sich um einen Filterport.

KAPITEL 5

Mapping-Optimierung

Dieses Kapitel umfasst die folgenden Themen:

- Mapping-Optimierung - Übersicht, 42
- Optimierungsmethoden, 43
- Pushdown-Optimierung, 50
- Lesen in einem einzigen Durchlauf, 52
- Filteroptimierung, 53
- Optimierung der Datentypkonvertierung, 53
- Fehler-Tracing, 54

Mapping-Optimierung - Übersicht

Optimieren Sie Mappings, damit der Datenintegrationsdienst Daten effizient umwandeln und verschieben kann. Das Implementieren der Optimierung auf Mapping-Ebene kann zeitaufwendig sein. Die Mapping-Leistung kann dadurch jedoch erheblich gesteigert werden.

Die Optimierungsaufgaben gelten für reguläre Mappings, Lese- und Schreib-Mappings für logische Datenobjekte, virtuelle Tabellen-Mappings und Operations-Mappings. Konzentrieren Sie sich auf die Optimierung auf Mapping-Ebene, nachdem Sie die Targets und Quellen optimiert haben.

Zum Optimieren eines Mappings können Sie die folgenden Aufgaben durchführen:

- Konfigurieren Sie das Mapping mit der geringstmöglichen Anzahl an Umwandlungen und Ausdrücken, um bestmögliche Ergebnisse zu erzielen.
- Löschen Sie nicht benötigte Verknüpfungen zwischen Umwandlungen, um die Menge der zu verschiebenden Daten zu minimieren.
- Wählen Sie eine Optimierungsebene zur Angabe der Optimierungsmethoden aus, die vom Datenintegrationsdienst auf das Mapping angewendet werden können. Wenn der Datenintegrationsdienst ein Mapping optimiert, versucht er, die Menge der zu verarbeitenden Daten zu verringern. Beispielsweise kann der Datenintegrationsdienst mithilfe der frühen Auswahloptimierung einen Filter näher an die Quelle verschieben. Mit der kostenbasierten Optimierungsmethode kann die Join-Verarbeitungsreihenfolge geändert werden.
- Wählen Sie einen Pushdown-Typ, damit der Datenintegrationsdienst ermitteln kann, ob er die Umwandlungslogik teilweise oder vollständig in die Quelldatenbank übertragen kann.
- Konfigurieren Sie die Zwischenspeicherung von Datenobjekten, um dem Datenintegrationsdienst beim Ausführen eines Mappings das Zwischenspeichern logischer Datenobjekte sowie den Zugriff auf vorgefertigte logische Datenobjekte zu ermöglichen. Standardmäßig extrahiert der

Datenintegrationsdienst Quelldaten und erstellt benötigte Datenobjekte, wenn er ein Mapping ausführt. Die Mapping-Leistung wird erhöht, wenn der Datenintegrationsdienst auf vorgefertigte Datenobjekte zugreifen kann.

- Geben Sie an, ob bei der SQL-Umwandlung, der Webdienst-Verbraucher-Umwandlung und der Java-Umwandlung Nebeneffekte auftreten, wenn Sie diese Umwandlungen konfigurieren. Bestimmte Umwandlungen haben Nebeneffekte, die die Optimierung einschränken. Beispielsweise kann bei einer Umwandlung ein Nebeneffekt auftreten, wenn die Umwandlung in eine Datei oder Datenbank schreibt, eine Anzahl erhöht, eine Ausnahme auslöst oder eine E-Mail schreibt. In den meisten Fällen erkennt der Datenintegrationsdienst die Umwandlungen, die Nebeneffekte haben und die Optimierung einschränken.

Verwandte Themen:

- "Datenobjekt-Caching" auf Seite 66

Optimierungsmethoden

Der Datenintegrationsdienst wendet Optimierungsmethoden an, um die Anzahl der in der Zuordnung zu verarbeitenden Zeilen zu verringern. Sie können die Optimierungsebene für die Zuordnung konfigurieren, um zu beschränken, welche Optimierungsmethoden der Datenintegrationsdienst anwendet.

Der Datenintegrationsdienst kann die folgenden Optimierungsmethoden anwenden:

- Pushdown-Optimierung
- Frühe Projektionsoptimierung
- Frühe Auswahloptimierung
- Verzweigungsbereinigungsoptimierung
- Push-Into-Optimierung
- Vorhersageoptimierung
- Globale Vorhersageoptimierung
- Kostenbasierte Optimierung
- Dataship-Join-Optimierung
- Semi-Join-Optimierung

Der Datenintegrationsdienst kann gleichzeitig mehrere Optimierungsmethoden auf ein Mapping anwenden. Der Datenintegrationsdienst wendet beispielsweise die Methoden zur frühen Projektionsoptimierung, zur Vorhersageoptimierung, zur globalen Vorhersageoptimierung, zur frühen Auswahloptimierung oder zur Push-Into-Optimierung an.

Optimierungsebenen

Der Datenintegrationsdienst optimiert Mappings auf der Grundlage der von Ihnen konfigurierten Optimierungsebene. Konfigurieren Sie die Optimierungsebene, wenn Sie möchten, dass das Mapping eine andere as die normale Optimierungsebene verwendet. Standardmäßig verwendet jedes Mapping die normale Optimierungsebene.

Sie können eine der folgenden Optimierungsebenen auswählen:

Keiner

Der Datenintegrationsdienst wendet keine Optimierung an.

Minimal

Der Datenintegrationsdienst wendet die frühe Projektionsoptimierungsmethode an.

Normal

Der Datenintegrationsdienst wendet folgende Optimierungsmethoden an: frühe Projektion, frühe Auswahl, Verzweigungsbereinigung, Push-Into, globale Vorhersage und Vorhersage. Normal ist die Standard-Optimierungsebene.

Komplett

Der Datenintegrationsdienst wendet folgende Optimierungsmethoden an: kostenbasiert, frühe Projektion, frühe Auswahl, Verzweigungsbereinigung, Vorhersage, Push-Into, Semi-Join und Dataship-Join.

Der Datenintegrationsdienst wendet die normale Optimierungsebene an, wenn Sie ein Mapping über das Menü **Ausführen** oder über den Zuordnungseditor im Developer Tool ausführen. Wenn Sie das Mapping über das Menü **Ausführen** ausführen, wendet der Datenintegrationsdienst die Optimierungsebene aus der Zuordnungskonfiguration an. Wenn Sie das Mapping über die Befehlszeile ausführen, wendet der Datenintegrationsdienst die Optimierungsebene aus den Zuordnungsbereitstellungs-Eigenschaften in der Anwendung an.

Hinweis: Der Datenintegrationsdienst weist die Pushdown-Optimierungsmethode nicht mit einer Optimierungsebene zu. Sie können die Pushdown-Optimierung für ein Mapping in den Laufzeiteigenschaften des Mappings konfigurieren.

Filteroptimierungen

Die Filteroptimierung erhöht die Leistung, indem die Anzahl der Zeilen, die durch das Mapping geleitet werden, verringert wird. Der Data Integration Service kann die "Early Selection"- oder die Push-Into-Optimierung anwenden.

Wenn der Data Integration Service eine Filteroptimierungsmethode anwendet, verschiebt er einen Filter so nahe wie möglich an die Quelle im Mapping. Wenn der Data Integration Service einen Filter nicht vor eine Umwandlung in einem Mapping verschieben kann, kann er die Filterlogik gegebenenfalls in eine Umwandlung leiten.

Frühe Projektionsoptimierungsmethode

Wenn der Datenintegrationsdienst die frühe Projektionsoptimierungsmethode anwendet, identifiziert er ungenutzte Ports und entfernt die Links zwischen diesen Ports.

Die frühe Projektionsoptimierungsmethode verbessert die Leistung, indem die Menge der Daten, die vom Datenintegrationsdienst zwischen Umwandlungen verschoben werden, verringert wird. Wenn der Datenintegrationsdienst ein Mapping verarbeitet, verschiebt er die Daten aus allen verbundenen Ports in einem Mapping von einer Umwandlung zu einer anderen Umwandlung. In großen komplexen Mappings oder in Mappings mit geschachtelten Mapplets stellen einige Ports möglicherweise keine Daten für das Target bereit. Der Datenintegrationsdienst identifiziert die Ports, die keine Daten für das Ziel bereitstellen. Nachdem der Datenintegrationsdienst ungenutzte Ports identifiziert hat, entfernt er die Links zwischen allen ungenutzten Ports aus dem Mapping.

Der Datenintegrationsdienst entfernt nicht alle Links. Beispielsweise entfernt er nicht die folgenden Links:

- Mit einer Umwandlung verbundene Verknüpfungen mit Nebeneffekten.
- Links, die mit Umwandlungen verbunden sind, die eine ABORT()- oder ERROR()-Funktion aufrufen, E-Mails senden oder eine gespeicherte Prozedur aufrufen.

Wenn der Datenintegrationsdienst bestimmt, dass alle Ports in einer Umwandlung nicht genutzt werden, entfernt er alle Umwandlungslinks außer dem Link zu dem Port mit den wenigsten Daten. Der Datenintegrationsdienst entfernt nicht die ungenutzte Umwandlung aus dem Mapping.

Diese Optimierungsmethode wird vom Developer Tool standardmäßig aktiviert.

Vorhersageoptimierungsmethode

Wenn der Datenintegrationsdienst die Vorhersageoptimierungsmethode anwendet, überprüft er die von einem Mapping generierten Vorhersage-Expressionen. Er bestimmt, ob er die Expressionen vereinfachen oder umschreiben kann, um die Mapping-Leistung zu erhöhen.

Wenn der Datenintegrationsdienst ein Mapping ausführt, generiert er Abfragen der Mapping-Quellen und führt basierend auf der Mapping-Logik und den Umwandlungen im Mapping Operationen für die Abfrageergebnisse aus. Die Abfragen und Operationen enthalten häufig Vorhersage-Expressionen. Vorhersage-Expressionen stellen die Bedingungen dar, die von den Daten erfüllt werden müssen. Die Filter- und die Join-Bedingung in der Filter- bzw. Joiner-Umwandlung sind Beispiele für Vorhersage-Expressionen.

Der Datenintegrationsdienst versucht mit der Vorhersageoptimierungsmethode außerdem, zu einem möglichst frühen Zeitpunkt im Mapping Vorhersage-Expressionen anzuwenden, um die Mapping-Leistung zu erhöhen.

Der Datenintegrationsdienst leitet aus vorhandenen Vorhersage-Expressionen Beziehungen ab und erstellt neue Vorhersage-Expressionen. Angenommen, ein Mapping enthält eine Joiner-Umwandlung mit der Join-Bedingung „A=B“ und eine Filter-Umwandlung mit der Filterbedingung „A>B“. Der Datenintegrationsdienst kann der Join-Bedingung „B>5“ hinzufügen.

Der Datenintegrationsdienst wendet die Vorhersageoptimierungsmethode mit der frühen Auswahloptimierungsmethode an, wenn er beide Methoden auf ein Mapping anwenden kann. Wenn der Datenintegrationsdienst beispielsweise neue Filterbedingungen über die Vorhersageoptimierungsmethode erstellt, versucht er außerdem, sie mit der frühen Auswahloptimierungsmethode im Mapping aufwärts zu verschieben. Die Anwendung beider Optimierungsmethoden bietet eine bessere Mapping-Leistung als die Anwendung nur einer dieser Methoden.

Der Datenintegrationsdienst wendet die Vorhersageoptimierungsmethode an, wenn dies die Leistung erhöht. Der Datenintegrationsdienst wendet diese Methode nicht an, wenn die Anwendung die Mapping-Ergebnisse ändert oder die Mapping-Leistung verringert. Der Datenintegrationsdienst wendet diese Optimierungsmethode standardmäßig an.

Regeln und Richtlinien für die Vorhersageoptimierung

Wenn der Datenintegrationsdienst eine Vorhersage-Expression neu schreibt, wendet er zum Optimieren der Expression mathematische Logik auf diese an.

Der Datenintegrationsdienst kann die folgenden Aktionen einzeln oder gemeinsam ausführen:

- Bestimmen von übereinstimmenden Variablen in den Vorhersage-Expressionen im Mapping und Generieren von vereinfachten Expressionen auf Grundlage der Übereinstimmungen.
- Bestimmen und Entfernen von redundanten Vorhersagen in den Expressionen im Mapping.
- Extrahieren von untergeordneten Expressionen aus disjunktiven Klauseln und Generieren von mehreren vereinfachten Expressionen auf Grundlage der untergeordneten Expressionen.
- Normalisieren einer Vorhersage-Expression.
- Anwenden von Vorhersage-Expressionen zu einem möglichst frühen Zeitpunkt im Mapping.

Der Datenintegrationsdienst wendet keine Vorhersageoptimierung auf ein Mapping an, wenn dieses Umwandlungen enthält, bei denen ein Datentypkonflikt zwischen verbundenen Ports auftritt.

Der Datenintegrationsdienst wendet keine Vorhersageoptimierung auf eine Umwandlung an, wenn eine der folgenden Bedingungen zutrifft:

- Die Umwandlung enthält explizite Standardwerte für verbundene Ports.
- Die Umwandlung hat Nebenwirkungen.
- Die Umwandlung lässt das Verschieben von Vorhersagen nicht zu. Beispielsweise hat eine Umwandlung, die Nebenwirkungen hat, diese Beschränkung.

Im Developer Tool ist die Vorhersageoptimierungsmethode standardmäßig aktiviert.

Kostenbasierte Optimierungsmethode

Mit kostenbasierter Optimierung wertet der Datenintegrationsdienst ein Mapping aus, generiert semantisch äquivalente Mappings und führt das Mapping mit der bestmöglichen Leistung aus. Kostenbasierte Optimierung verringert die Laufzeit für Mappings, die Operationen für benachbarte innere und vollständig äußere Joins durchführen.

Semantisch äquivalente Mappings führen identische Funktionen aus und erzeugen die gleichen Ergebnisse. Zum Generieren von semantisch äquivalenten Mappings unterteilt der Datenintegrationsdienst das ursprüngliche Mapping in Fragmente. Anschließend bestimmt der Datenintegrationsdienst, welche Mapping-Fragmente optimiert werden können.

Der Datenintegrationsdienst kann während der Optimierung Umwandlungen in einem Fragment hinzufügen, entfernen oder neu anordnen. Der Datenintegrationsdienst überprüft, ob die optimierten Fragmente die gleichen Ergebnisse wie die ursprünglichen Fragmente erzeugen, und bildet alternative Mappings, die die optimierten Fragmente nutzen.

Der Datenintegrationsdienst kann auch ein Join mit sortierter Zusammenführung anwenden, wenn er feststellt, dass die Leistung bei einem Join mit sortierter Zusammenführung besser ist als bei einem Join mit geschachtelter Schleife. Ein Join mit sortierter Zusammenführung verwendet eine Sortierreihenfolge, um zwei Datensätze vor dem Join anzuordnen. Ein Join mit geschachtelter Schleife verwendet geschachtelte Schleifen zum Verbinden von zwei Datensätzen. Der Datenintegrationsdienst verwendet möglicherweise die Sortierinformationen in den Quellen oder erstellt eine Sortiererumwandlung, wenn die Kosten zum Sortieren der Daten geringer sind als beim Verarbeiten eines Join mit geschachtelter Schleife.

Der Datenintegrationsdienst generiert alle oder fast alle Mappings, die mit dem ursprünglichen Mapping semantisch äquivalent sind. Er verwendet Profiling-Statistiken oder Datenbankstatistik, um die Kosten für das Original-Mapping und jedes alternative Mapping zu berechnen. Anschließend ermittelt er das Mapping, das am schnellsten ausgeführt wird. Der Datenintegrationsdienst führt eine Validierungsprüfung zum Ermitteln des besten alternativen Mappings aus, um sicherzustellen, dass es gültig ist und die gleichen Ergebnisse wie das ursprüngliche Mapping erzeugt.

Der Datenintegrationsdienst zwischenspeichert das beste alternative Mapping. Wenn Sie ein Mapping ausführen, ruft der Datenintegrationsdienst das alternative Mapping ab und führt es anstelle des ursprünglichen Mappings aus.

Im Developer Tool wird diese Methode nicht standardmäßig aktiviert.

Dataship-Join-Optimierungsmethode

Mit der Dataship-Join-Optimierungsmethode wird versucht, kleinere Datensätze neben größeren Datensätze zu platzieren, um die Join-Bearbeitungszeit zu verringern. Der Datenintegrationsdienst versucht, die Dataship-Join-Optimierungsmethode anzuwenden, wenn zwischen zwei Tabellen bedeutende Größenunterschiede bestehen.

Der Datenintegrationsdienst kann beispielsweise die Dataship-Join-Optimierungsmethode anwenden, um eine 10.000 Zeilen umfassende Master-Tabelle mit einer 1.000.000 Zeilen umfassenden Detail-Tabelle zu

verbinden. Um Dataship-Join durchzuführen, erstellt der Datenintegrationsdienst eine temporäre Staging-Tabelle in der Datenbank, die die größere Detail-Tabelle enthält. Dann kopiert der Datenintegrationsdienst die kleinere Master-Tabelle in eine temporäre Tabelle und verknüpft die Daten in der temporären Tabelle mit den Daten in der größeren Detailtabelle. Nachdem der Datenintegrationsdienst die Join-Operation durchgeführt hat, wird die Joiner-Umwandlungslogik in der Datenbank verarbeitet.

Bevor der Datenintegrationsdienst die Dataship-Join-Optimierungsmethode anwendet, führt er Analysen aus, um zu bestimmen, ob eine Dataship-Join-Optimierung möglich ist und den Aufwand lohnt. Wenn die Analysen ergeben, dass die Leistung durch diese Methode wahrscheinlich erhöht wird, wird sie vom Datenintegrationsdienst auf die Zuordnung angewendet. Anschließend analysiert der Datenintegrationsdienst die Zuordnung erneut, um zu bestimmen, ob weitere Möglichkeiten für Dataship-Join-Optimierung bestehen. Ggf. führt er weitere Optimierungen aus.

Im Developer Tool wird diese Methode nicht standardmäßig aktiviert.

Anforderungen für die Dataship-Join-Optimierung zur Erhöhung der Leistung

Die Dataship-Join-Optimierungsmethode erhöht nicht immer die Leistung. Die folgenden Faktoren wirken sich bei der Dataship-Join-Optimierung auf die Mapping-Leistung aus:

- Die Master-Quelle der Joiner-Umwandlung muss wesentlich weniger Zeilen als die Detailquelle enthalten.
- Die Detailquelle muss groß genug sein, damit sich die Optimierung lohnt. Wenn die Detailquelle nicht groß genug ist, stellt der Datenintegrationsdienst fest, dass das Lesen aller Daten aus der Master- und Detail-Quelle schneller ohne das Anwenden der Dataship-Join-Optimierungsmethode erfolgt.

Regeln und Richtlinien für die Dataship-Join-Optimierung

Der Datenintegrationsdienst kann Dataship-Join-Optimierung auf eine Joiner-Umwandlung anwenden, wenn diese die folgenden Anforderungen erfüllt:

- Der Join-Typ muss „Normal“, „Master-Outer“ oder „Detail-Outer“ lauten.
- Die Detail-Pipeline muss aus einer relationalen Datenquelle stammen.
- Wenn die Zuordnung zielbasierte Commits verwendet, muss der Bereich der Joiner-Umwandlung „Alle Eingaben“ lauten.
- Die Master-Pipeline und die Detail-Pipeline dürfen nicht Informationen gemeinsam nutzen.
- Die Zuordnung darf keine Verzweigung zwischen der Detailquelle und der Joiner-Umwandlung enthalten.
- Der Datenintegrationsdienst kann die Dataship-Join-Optimierungsmethode nicht anwenden, wenn die Datenbank mit der Detailseite dieses Joins eine IBM DB2-Datenbank ist, die Unicode-Codierung nicht unterstützt.

Semi-Join Optimierungsmethode

Die Semi-Join Optimierungsmethode ändert die Join-Operationen im Mapping, um die Menge der aus der Quelle extrahierten Daten zu verringern.

Der Datenintegrationsdienst wendet die Semi-Join Optimierungsmethode auf eine Joiner-Umwandlung an, wenn eine Eingabegruppe über mehr Zeilen als die andere Gruppe verfügt und wenn die größere Gruppen viele Zeilen enthält, für die in der kleineren Gruppen keine Übereinstimmung gemäß der Join-Bedingung vorhanden ist. Der Datenintegrationsdienst versucht, die Größe der Datenmenge eines Join-Operanden zu verringern, indem er die Zeilen aus der kleineren Gruppe liest, die übereinstimmenden Zeilen in der größeren Gruppe sucht und dann die Join-Operation durchführt. Das Verringern der Größe der Datenmenge verbessert die Mapping-Leistung, da der Datenintegrationsdienst nicht mehr unnötige Zeilen aus der größeren Quellgruppe liest. Der Datenintegrationsdienst verschiebt die Join-Bedingung in die größere Quellgruppe und liest nur die Zeilen, die mit der kleineren Gruppe übereinstimmen.

Bevor der Datenintegrationsdienst die Semi-Join Optimierungsmethode anwendet, führt er Analysen aus, um zu bestimmen, ob eine Dataship-Join-Optimierung möglich ist und den Aufwand lohnt. Wenn die Analysen ergeben, dass die Leistung durch diese Methode wahrscheinlich erhöht wird, wird sie vom Datenintegrationsdienst auf das Mapping angewendet. Anschließend analysiert der Datenintegrationsdienst das Mapping erneut, um zu bestimmen, ob weitere Möglichkeiten für Semi-Join Optimierung bestehen. Ggf. führt er weitere Optimierungen aus.

Im Developer Tool wird diese Methode nicht standardmäßig aktiviert.

Anforderungen für die Semi-Join-Optimierung zur Erhöhung der Leistung

Die Semi-Join Optimierungsmethode erhöht nicht immer die Leistung. Die folgenden Faktoren wirken sich bei der Semi-Join-Optimierung auf die Mapping-Leistung aus:

- Die Master-Quelle der Joiner-Umwandlung muss wesentlich weniger Zeilen als die Detailquelle enthalten.
- Die Detailquelle muss groß genug sein, damit sich eine Optimierung lohnt. Wenn der Data Integration Service die Semi-Join Optimierung anwendet, erhöht sich die Dauer der Mapping-Verarbeitung. Wenn die Detailquelle klein ist, überschreitet die zum Anwenden der Semi-Join-Methode erforderliche Zeit möglicherweise die Zeit, die zum Verarbeiten aller Zeilen in der Detailquelle benötigt wird.
- Der Data Integration Service muss für eine Joiner-Umwandlung eine Statistik zur Anzahl der Quellzeilen abrufen können, um die zeitlichen Anforderungen der regulären Join-Operation mit den zeitlichen Anforderungen der Semi-Join-Operation vergleichen zu können.

Regeln und Richtlinien für die Semi-Join-Optimierung

Der Data Integration Service kann Semi-Join-Optimierung auf eine Joiner-Umwandlung anwenden, wenn diese die folgenden Anforderungen erfüllt:

- Der Join-Typ muss „Normal“, „Master-Outer“ oder „Detail-Outer“ lauten. Die Joiner-Umwandlung kann keinen Full Outer Join ausführen.
- Die Detail-Pipeline muss aus einer relationalen Datenquelle stammen.
- Die Join-Bedindung muss eine gültige Sort-Merge-Join-Bedingung sein. Das heißt, dass jede Klausel einem Master-Port und einem Detail-Port entsprechen muss. Wenn mehrere Klauseln vorhanden sind, müssen sie durch AND verknüpft werden.
- Wenn das Mapping keine Target-basierten Commits verwendet, muss der Bereich der Joiner-Umwandlung „Alle Eingaben“ lauten.
- Die Master-Pipeline und die Detail-Pipeline dürfen nicht Informationen gemeinsam nutzen.
- Das Mapping darf keine Verzweigung zwischen der Detailquelle und der Joiner-Umwandlung enthalten.

Frühe Auswahloptimierungsmethode

Wenn der Datenintegrationsdienst die frühe Auswahloptimierungsmethode anwendet, werden die Filterumwandlungen in einer Zuordnung aufgeteilt, verschoben oder entfernt. Filter werden in der Zuordnung aufwärts näher zur Quelle verschoben.

Der Data Integration Service kann eine Filter-Umwandlung aufteilen, wenn die Filterbedingung eine Konjunktion ist. Der Data Integration Service kann beispielsweise die Filterbedingung „A>100 AND B<50“ in die beiden einfacheren Bedingungen „A>100“ und „B<50“ aufteilen. Wenn der Datenintegrationsdienst einen Filter aufteilt, verschiebt er die vereinfachten Filter in der Zuordnungs-Pipeline aufwärts näher zur Quelle. Der Data Integration Service verschiebt die Filter getrennt in der Pipeline aufwärts, wenn er den Filter aufteilt.

Die Methode zur frühen Auswahloptimierung ist standardmäßig aktiviert, wenn Sie die normale oder vollständige Optimierungsebene im Developer Tool auswählen. Der Datenintegrationsdienst ignoriert die

frühe Auswahloptimierung, wenn eine Umwandlung, die vor der Filterumwandlung stattfindet, Nebeneffekte erzeugt. Der Datenintegrationsdienst kann nicht feststellen, ob die SQL-, Webdienstbenutzer- und Java-Umwandlung Nebeneffekte mit sich bringen. Sie können die frühe Auswahloptimierung für diese Umwandlungen konfigurieren, wenn sie keine Nebeneffekte erzeugen.

Sie können die frühe Auswahloptimierung deaktivieren, wenn sie die Leistung nicht erhöht. Diese Optimierungsmethode wird vom Datenintegrationsdienst standardmäßig aktiviert.

Globale Vorhersageoptimierungsmethode

Wenn der Datenintegrationsdienst die globale Vorhersageoptimierungsmethode verwendet, entfernt er Zeilen, die ausgefiltert werden können, so früh wie möglich im Mapping. Dadurch wird der Umfang der Zeilen reduziert, die für das Mapping verarbeitet werden müssen. Die globale Vorhersageoptimierungsmethode umfasst sowohl die Methode zur Vorhersageoptimierung als auch zur frühen Auswahloptimierung.

Angenommen, ein Mapping enthält eine Joiner-Umwandlung mit der Join-Bedingung „A=B“ und eine Filter-Umwandlung mit der Filterbedingung „A>5“. Der Datenintegrationsdienst kann möglicherweise „B>5“ zur Join-Bedingung hinzufügen und die Filterumwandlung näher an die Quelle verschieben.

Die globale Vorhersageoptimierungsmethode setzt Vorhersageausdrücke effektiver ein als die Vorhersageoptimierungsmethode. Die globale Vorhersageoptimierungsmethode bestimmt, ob sie die Ausdrücke vereinfachen oder neu schreiben kann, um die Mapping-Leistung zu erhöhen. Sie versucht zudem, Vorhersageausdrücke im Mapping so früh wie möglich anzuwenden, um die Mapping-Leistung zu verbessern.

Die globale Vorhersageoptimierungsmethode leitet Filter ab und überträgt sie näher an die Quelle, wenn das Mapping verschachtelte Joiner oder Verzweigungen mit Filtern in jeder Verzweigung enthält. Wenn der Datenintegrationsdienst die globale Vorhersageoptimierungsmethode verwendet, teilt er die auf Filter auf, die sich näher an der Quelle befinden, oder entfernt Filter in einem Mapping.

Methode zur Optimierung der Verzweigungsbereinigung

Der Datenintegrationsdienst kann die Methode zur Optimierung der Verzweigungsbereinigung auf Umwandlungen anwenden, die keine Zeilen zum Ziel in einem Mapping beitragen.

Der Datenintegrationsdienst kann eine Filterumwandlung unter Umständen entfernen, wenn die Filterbedingung für die Datenzeilen mit FALSE bewertet wird. Beispiel: Ein Mapping besteht aus zwei Filterumwandlungen, die Daten aus zwei relationalen Quellen filtern. Eine Filterumwandlung weist die Filterbedingung Country=US und die andere Filterumwandlung die Filterbedingung Country=Canada auf. Eine Unionumwandlung verknüpft die beiden relationalen Quellen und weist die Filterbedingung Country=US auf. Der Datenintegrationsdienst entfernt die Filterumwandlung mit der Filterbedingung Country=Canada unter Umständen aus dem Mapping.

Das Developer-Tool aktiviert standardmäßig die Methode zur Optimierung der Verzweigungsbereinigung, wenn Sie eine normale oder vollständige Optimierungsebene auswählen. Sie können die Verzweigungsbereinigung deaktivieren, wenn die Leistung durch die Optimierung nicht erhöht wird, indem Sie die Optimierungsebene auf „Minimal“ oder „Keine“ setzen.

Push-Into-OptimierungMethode

Mit der push-into-Optimierung überträgt der Datenintegrationsdienst die Filter-Umwandlungslogik in die Umwandlung, die sich direkt oberhalb der Filterumwandlung im Mapping befindet. Die Push-Into-Optimierung erhöht die Leistung, indem sie die Anzahl der Zeilen reduziert, die das Mapping durchlaufen.

Der Datenintegrationsdienst überträgt keine Filterlogik in eine andere Umwandlung, wenn die Umwandlung Nebenwirkungen hat. Der Datenintegrationsdienst kann nicht feststellen, ob die SQL-Umwandlung, Webdienstverbraucher-Umwandlung und Java-Umwandlung Nebenwirkungen hat. Sie können jedoch die SQL-

Umwandlung, die Webdienstverbraucher-Umwandlung und die Java-Umwandlung für die Push-into-Optimierung konfigurieren.

Pushdown-Optimierung

Wenn der Datenintegrationsdienst Pushdown-Optimierung anwendet, überträgt er die Umwandlungslogik in die Quelldatenbank. Der Datenintegrationsdienst übersetzt die Umwandlungslogik in SQL-Abfragen und sendet diese an die Datenbank. Die Quelldatenbank führt die SQL-Abfragen aus, um die Umwandlungen zu verarbeiten.

Die Pushdown-Optimierung erhöht die Mapping-Leistung, wenn die Quelldatenbank die Umwandlungslogik schneller verarbeiten kann als der Datenintegrationsdienst. Der Datenintegrationsdienst liest außerdem weniger Daten aus der Quelle.

Der Betrag an Umwandlungslogik, den der Datenintegrationsdienst in die Quelldatenbank überträgt, hängt von der Datenbank, der Umwandlungslogik und der Mapping-Konfiguration ab. Der Datenintegrationsdienst verarbeitet sämtliche Umwandlungslogik, die er nicht in die Datenbank übertragen kann.

Wenn Sie Pushdown-Optimierung anwenden, analysiert der Datenintegrationsdienst das optimierte Mapping von der Quelle bis zum Target oder bis er eine nachfolgende Umwandlung erreicht, die er nicht in die Quelldatenbank übertragen kann. Der Datenintegrationsdienst generiert für jede Quelle, in die Umwandlungslogik übertragen wurde, eine SELECT-Abfrage und führt diese aus. Der Datenintegrationsdienst kann auch eine INSERT-Abfrage generieren, wenn das Ziel in die Datenbank übertragen wurde. Der Datenintegrationsdienst liest die Ergebnisse der SQL-Abfragen und verarbeitet die übrigen Umwandlungen im Mapping.

Der Datenintegrationsdienst wendet Pushdown-Optimierung auf ein Mapping an, wenn Sie in den Laufzeiteigenschaften des Mappings den Pushdown-Typ auswählen.

Sie können die folgenden Pushdown-Typen auswählen:

- Keine. Wählt keinen Pushdown-Typ für das Mapping aus.
- Quelle. Der Datenintegrationsdienst versucht, möglichst viel Umwandlungslogik in die Quelldatenbank zu übertragen.
- Vollständig. Der Datenintegrationsdienst überträgt die vollständige Umwandlungslogik in die Quelldatenbank.

Sie können auch einen Zeichenfolgenparameter für den Pushdown-Typ erstellen und die folgenden Parameterwerte verwenden:

- Keiner
- Quelle
- Vollständig

Vollständige Pushdown-Optimierung

Wenn der Datenintegrationsdienst die vollständige Pushdown-Optimierung anwendet, überträgt er die gesamte Umwandlungslogik in der Zuordnung in die Quelldatenbank. Sie können die vollständige Übertragung in den Laufzeiteigenschaften der Zuordnung konfigurieren.

Vollständige Pushdown-Optimierung ist am besten geeignet, wenn sich Quelle und Ziel in derselben Datenbank befinden oder wenn Umwandlungen wie Aggregator- und Filterumwandlungen in der Quelldatenbank verarbeitet werden und die Menge der zu verschiebenden Daten minimiert wird. Wenn eine

Zuordnung beispielsweise eine Teradaten-Quelle und ein Teradaten-Ziel enthält, müssen Sie vollständige Pushdown-Optimierung konfigurieren, um die gesamte Umwandlungslogik für die Verarbeitung von einer Teradaten-Quelldatenbank in eine Teradaten-Zieldatenbank zu übertragen.

Wenn Sie eine Zuordnung mit einer Update-Strategieumwandlung oder vollständigen Übertragung konfigurieren, müssen Sie die Pushdown-Kompatibilität für die Zuordnung festlegen.

Der Datenintegrationsdienst kann eine Zuordnung mit einer Update-Strategieumwandlung in den folgenden Szenarien übertragen:

- Wenn die mit der Updatestrategie-Umwandlung verbundene Zielumwandlung mehrere Zeilen empfängt, die nicht denselben Schlüssel aufweisen.
- Wenn die mit der Updatestrategie-Umwandlung verbundene Zielumwandlung mehrere Zeilen mit demselben Schlüssel empfängt, die neu sortiert werden können.

Der Datenintegrationsdienst kann eine Zuordnung mit einer Update-Strategieumwandlung im folgenden Szenario nicht übertragen:

- Wenn die mit der Updatestrategie-Umwandlung verbundene Zielumwandlung mehrere Zeilen mit demselben Schlüssel empfängt, die nicht neu sortiert werden können.

Sie können auch einen Pushdown-Kompatibilitätsparameter in der Zuordnung verwenden. Sie können die folgenden Parameterwerte verwenden:

- noMultipleRowsWithSameKeyOnTarget
- reorderAllowedForMultipleRowsWithSameKey
- reorderNotAllowedForRowsWithSameKey

Der Datenintegrationsdienst kann die vollständige Pushdown-Optimierung für die folgenden Quellen verwenden:

- Oracle
- IBM DB2
- Microsoft SQL Server
- Teradata
- Netezza
- Greenplum
- SAP HANA

Quell-Pushdown

Wenn der Datenintegrationsdienst ein Quell-Pushdown anwendet, analysiert er das Mapping von der Quelle bis zum Ziel oder bis er auf eine nachgeordnete Umwandlung trifft, die sich nicht in die Quelldatenbank übertragen lässt.

Der Datenintegrationsdienst generiert und führt eine SELECT-Anweisung basierend auf der Umwandlungslogik für jede Umwandlung aus, die er in die Datenbank verschieben kann. Anschließend liest er die Ergebnisse dieser SQL-Abfrage aus und verarbeitet die restlichen Umwandlungen.

Sie können ein Mapping so konfigurieren, dass es Quell-Pushdown verwendet, wenn die Quelle und das Ziel in unterschiedlichen Datenbanken vorliegen. Wenn ein Mapping beispielsweise eine Teradaten-Quelle und ein Oracle-Ziel enthält, können Sie das Quell-Pushdown so konfigurieren, dass ein Teil der Umwandlungslogik zur Verarbeitung in die Teradaten-Zieldatenbank übertragen wird.

Regeln und Richtlinien für die Pushdown-Optimierung

Der Datenintegrationsdienst kann Umwandlungslogik in die Quelldatenbank übertragen.

Für die Pushdown-Optimierung gelten die folgenden Regeln und Richtlinien:

- Der Datenintegrationsdienst kann Lookup- und Joiner-Umwandlungslogik in die Quelldatenbank übertragen, wenn sich die Quellen in demselben Datenbankverwaltungssystem befinden und dieselben Verbindungen verwenden.
- Der Datenintegrationsdienst kann Umwandlungslogik nicht in eine Quelle mit binärem Datentyp übertragen.
- Der Datenintegrationsdienst deaktiviert die Pushdown-Optimierung, wenn Sie über eine IBM DB2-Datenquelle verfügen und die Spaltenpräzision beim Dezimal-Datentyp zwischen 28 und 31 Stellen beträgt.
- Der Datenintegrationsdienst aktiviert die Pushdown-Optimierung bei einem SQL-Datendienst oder einem Webdienst standardmäßig. Sie können die Pushdown-Optimierung bei einem SQL-Datendienst oder Webdienst nicht deaktivieren.
- Der Datenintegrationsdienst kann keine Aggregator-Umwandlung, die einen Audruck mit Aggregatfunktionen und Nicht-Aggregatfunktionen enthält, in einen Port verschieben, der nicht gruppiert ist.

Lesen in einem einzigen Durchlauf

Das Lesen in einem einzigen Durchlauf ermöglicht es Ihnen, mehrere Targets mit einem konfigurierten Datenobjekt zu füllen. Verwenden Sie Lesen in einem einzigen Durchlauf, wenn mehrere Mappings die gleichen Quellen nutzen.

Zum Beheben von Engpässen beim Lesen in einem einzigen Durchlauf können Sie die folgenden Lösungen verwenden:

Fassen Sie die Umwandlungslogik für die einzelnen Mappings in einem Mapping zusammen und verwenden Sie für jede Quelle ein konfiguriertes Datenobjekt.

Der Data Integration Service liest jede Quelle ein Mal und sendet dann die Daten in verschiedenen Pipelines. Eine Zeile kann von allen Pipelines, einer beliebigen Kombination von Pipelines oder keinen Pipelines verwendet werden.

Angenommen, Sie verwenden eine Quelltabelle „Einkauf“, um täglich ein Aggregation und eine Rangliste zu erstellen. Wenn Sie die Aggregat-Umwandlung und die Rang-Umwandlung in separate Mappings einfügen, muss der Data Integration Service dieselbe Quelltabelle zweimal lesen. Wenn Sie jedoch die Aggregations- und Ranglogik in ein Mapping mit einem Quellqualifikator einfügen, liest der Data Integration Service die Tabelle „Einkauf“ einmal und sendet dann die entsprechenden Daten an die separaten Pipelines.

Klammern Sie gemeinsame Funktionen aus Mappings aus.

Wenn Sie Mappings ändern, um das Lesen in einem einzigen Durchlauf zu nutzen, können Sie diese Funktion optimieren, indem Sie gemeinsame Funktionen aus Mappings ausklammern. Wenn Sie z. B. sowohl für die Aggregat-Umwandlung als auch für die Rang-Umwandlung einen Prozentwert aus den Ports „Preis“ subtrahieren müssen, können Sie die Arbeit verringern, indem Sie den Prozentwert vor dem Aufteilen der Pipeline subtrahieren. Sie können den Prozentwert mit einer Expressions-Umwandlung subtrahieren und dann das Mapping nach der Umwandlung aufteilen.

Filteroptimierung

Sie können Mappings optimieren, indem Sie im benutzerdefinierten Datenobjekt filtern und Filter an einer frühen Stelle im Mapping platzieren.

Zum Beheben von Filterengpässen können Sie die folgenden Lösungen verwenden:

Verwenden Sie einen Filter in einem benutzerdefinierten Datenobjekt, um die Zeilen an der Quelle zu entfernen.

Wenn Sie Zeilen aus dem Mapping filtern, können Sie durch frühes Filtern im Datenfluss die Effizienz erhöhen. Verwenden Sie einen Filter in einem benutzerdefinierten Datenobjekt, um die Zeilen an der Quelle zu entfernen. Das benutzerdefinierte Datenobjekt begrenzt den aus einer relationalen Quelle extrahierten Satz von Zeilen.

Wenn Sie im benutzerdefinierten Datenobjekt keinen Filter verwenden können, verwenden Sie eine Filterumwandlung und platzieren Sie sie möglichst nahe am benutzerdefinierten Datenobjekt, um unnötige Daten früh im Datenfluss zu entfernen. Die benutzerdefinierte Datenumwandlung begrenzt den an ein Target gesendeten Satz von Zeilen.

Verwenden Sie in einer Aktualisierungsstrategie-Umwandlung einen Filter, wenn zurückgewiesene Zeilen nicht gespeichert werden müssen.

Zum Verbessern der Mapping-Leistung können Sie mit Hilfe einer Filter-Umwandlung zurückgewiesene Zeilen aus einer Aktualisierungsstrategie-Umwandlung löschen, wenn Sie zurückgewiesene Zeilen nicht speichern müssen.

Vermeiden Sie in Filter-Expressionen komplexe Expressionen.

Verwenden Sie in Filter-Expressionen keine komplexen Expressionen. Verwenden Sie zum Optimieren von Filter-Umwandlungen einfache ganzzahlige oder true/false-Expressionen.

Die Filter-Umwandlung filtert Daten in einem Mapping. Die Filter-Umwandlung filtert Zeilen aus jedem Quellentyp. Das benutzerdefinierte Datenobjekt filtert Zeilen aus relationalen Quellen. Die Filter-Umwandlung filtert Zeilen aus jedem Quellentyp.

Optimierung der Datentypkonvertierung

Sie können die Leistung erhöhen, indem Sie unnötige Datentypkonvertierungen vermeiden. Wenn z. B. ein Mapping Daten aus einer Ganzzahlspalte in eine Dezimalzahlspalte und dann zurück in eine Ganzzahlspalte verschiebt, führt die unnötige Datentypkonvertierung zu einer geringeren Leistung. Entfernen Sie nach Möglichkeit unnötige Datentypkonvertierungen aus Mappings.

Zum Beheben von Engpässen bei der Datentypkonvertierung können Sie die folgenden Lösungen verwenden:

Verwenden Sie statt anderer Datentypen ganzzahlige Werte, wenn Sie mit Lookup- und Filter-Umwandlungen Vergleiche ausführen.

Beispielsweise werden in vielen Datenbanken Postleitzahlen mit den Datentypen Char oder Varchar gespeichert. Wenn Sie die Postleitzahlen in den Datentyp Integer konvertieren, speichert die Lookup-Datenbank die US-Postleitzahl 94303-1234 als 943031234. Dies ermöglicht schnellere Lookup-Vergleiche anhand der Postleitzahl.

Konvertieren Sie die Quell-Datumsangaben mit Port-zu-Port-Konvertierungen in Strings, um die Mapping-Leistung zu erhöhen.

Sie können die Ports in Targets als Strings beibehalten oder die Ports in Date/Time-Ports ändern.

Fehler-Tracing

Um die Leistung zu verbessern, reduzieren Sie die Anzahl der Log-Ereignisse, die der Data Integration Service beim Ausführen des Mappings generiert. Verbessern Sie die Mapping-Leistung, indem Sie die Mapping-Optimierungsebene in der Mapping-Konfiguration oder in den Mapping-Bereitstellungs-Eigenschaften aktualisieren. Optimieren Sie Mappings mit Hilfe der kostenbasierten Optimierungsmethode.

Zum Beheben von Engpässen beim Fehler-Tracing können Sie die folgenden Lösungen verwenden:

Legen Sie in den Mapping-Eigenschaften die Tracingebene auf „Kurz“ fest.

Wenn ein Mapping eine große Anzahl von Umwandlungsfehlern enthält und Sie diese nicht korrigieren müssen, legen Sie in den Mapping-Eigenschaften die Tracingebene auf „Kurz“ fest. Auf dieser Tracingebene schreibt der Data Integration Service keine Fehlermeldungen oder Informationen auf Zeilenebene für zurückgewiesene Daten.

Wenn Sie das Mapping debuggen müssen und Sie die Tracingebene auf „Verbose“ festlegen, kann beim Ausführen des Mappings eine erhebliche Leistungseinbuße auftreten. Verwenden Sie nicht die Tracingebene „Verbose“, wenn Sie die Leistung optimieren. Die Tracingebene des Mappings überschreibt alle umwandlungsspezifischen Tracingebenen im Mapping. Dies wird nicht als langfristige Reaktion auf Umwandlungsfehler hoher Ebene empfohlen.

Ändern Sie die Optimierungsebene für das Mapping.

Wenn Sie feststellen, dass das Ausführen eines Mappings übermäßig viel Zeit erfordert, können Sie die Optimierungsebene für das Mapping ändern. Die Optimierungsebene bestimmt, welche Optimierungsmethoden der Data Integration Service zur Laufzeit auf das Mapping anwendet.

Sie legen die Optimierungsebene für ein Mapping in der Mapping-Konfiguration oder in den Mapping-Bereitstellungs-Eigenschaften fest. Der Data Integration Service wendet abhängig davon, wie das Mapping ausgeführt wird, unterschiedliche Optimierungsebenen auf das Mapping an.

Verwenden Sie die kostenbasierte Optimierungsmethode.

Die kostenbasierte Optimierungsmethode bewirkt, dass der Data Integration Service ein Mapping evaluiert, semantisch äquivalente Mappings generiert und das Mapping mit der besten Leistung ausführt. Diese Methode ist für Mappings besonders effektiv, die mehrere Join-Umwandlungen enthalten. Sie verringert die Laufzeit von Mappings, die benachbarte unsortierte Inner-Join-Vorgänge ausführen.

Semantisch äquivalente Mappings führen identische Funktionen aus und erzeugen die gleichen Ergebnisse. Zum Generieren von semantisch äquivalenten Mappings unterteilt der Data Integration Service das ursprüngliche Mapping in Fragmente. Anschließend bestimmt der Data Integration Service, welche Mapping-Fragmente optimiert werden können.

KAPITEL 6

Optimierung durch partitionierte Mappings

Dieses Kapitel umfasst die folgenden Themen:

- Optimierung durch partitionierte Mappings - Übersicht, 55
- Verwenden von mehreren CPUs, 56
- Erhöhen des maximalen Parallelismuswerts, 56
- Optimieren von Einfachdateien für Partitionierung, 57
- Optimieren von relationalen Datenbanken für Partitionierung, 58
- Optimieren von Umwandlungen für Partitionierung, 59

Optimierung durch partitionierte Mappings - Übersicht

Wenn Sie über die Partitionierungsoption verfügen, können Sie den Datenintegrationsdienst aktivieren, um Parallelismus beim Ausführen von Zuordnungen zu maximieren. Beim Maximieren des Parallelismus unterteilt der Datenintegrationsdienst die zugrunde liegenden Daten dynamisch in Partitionen und verarbeitet alle Partitionen gleichzeitig.

Wenn Mappings umfangreiche Datasets verarbeiten oder Umwandlungen enthalten, die komplizierte Berechnungen durchführen, kann die Mapping-Verarbeitung viel Zeit in Anspruch nehmen und einen geringen Datendurchsatz zur Folge haben. Wenn Sie Partitionierung für diese Mappings aktivieren, verwendet der Datenintegrationsdienst zusätzliche Threads zum Verarbeiten des Mappings.

Sie können die Leistung von partitionierten Mappings optimieren, indem Sie die folgenden Aufgaben durchführen:

- Verwenden Sie mehrere CPUs auf den Knoten, die die Mappings ausführen.
- Erhöhen Sie den maximalen Parallelismuswert für den Datenintegrationsdienst.
- Konfigurieren Sie Eigenschaften für Einfachdatei-Datenobjekte.
- Konfigurieren Sie relationale Datenbanken für die Optimierung der Partitionierung.
- Konfigurieren Sie Eigenschaften für Umwandlungen.

Verwenden von mehreren CPUs

Durch eine Erhöhung der Verarbeitungs-Threads werden die Knoten, die die Mappings ausführen, stärker belastet. Wenn die Knoten eine hohe CPU-Bandbreite aufweisen, kann die gleichzeitige Ausführung von Datenzeilen in einem Mapping die Mapping-Leistung optimieren.

Der Datenintegrationsdienst kann mehrere CPUs für die Verarbeitung eines Mappings mit mehreren Partitionen in Anspruch nehmen. Die Anzahl der CPUs, die der Dienst verwendet, hängt von Faktoren wie der Anzahl der Partitionierungspunkte, der Anzahl der pro Pipeline-Abschnitt erzeugten Threads und der für die Verarbeitung des Mappings benötigten Anzahl an Ressourcen ab. Ein einfaches Mapping wird in zwei Partitionen schneller ausgeführt, erfordert aber in der Regel die zweifache Menge an CPUs im Vergleich zur Ausführung des Mappings in einer einzelnen Partition.

Erhöhen des maximalen Parallelismuswerts

Maximaler Parallelismus bestimmt die maximale Zahl paralleler Threads, die einen einzelnen Pipeline-Abschnitt verarbeiten können. Konfigurieren Sie die Eigenschaft **Maximaler Parallelismus** für den Datenintegrationsdienst basierend auf den verfügbaren Hardware-Ressourcen. Sie können möglicherweise die Verarbeitungsdauer verringern, indem Sie den Wert für den maximalen Parallelismus erhöhen.

Beachten Sie die folgenden Richtlinien, wenn Sie den Wert für den maximalen Parallelismus erhöhen:

Erhöhen Sie den Wert basierend auf der Anzahl der verfügbaren CPUs.

Erhöhen Sie den Wert für den maximalen Parallelismus basierend auf der Anzahl der CPUs, die auf den Knoten verfügbar sind, auf denen Mappings ausgeführt werden. Wenn Sie den Wert für den maximalen Parallelismus erhöhen, verwendet der Datenintegrationsdienst mehr Threads zur Ausführung von Mappings und nutzt mehr CPUs. Eine einfache Zuordnung wird in zwei Partitionen schneller ausgeführt, erfordert aber in der Regel die zweifache Menge an CPUs im Vergleich zur Ausführung der Zuordnung in einer einzelnen Partition.

Beachten Sie die Gesamtanzahl der Verarbeitungs-Threads.

Beachten Sie die Gesamtanzahl der Verarbeitungs-Threads, wenn Sie den Wert für den maximalen Parallelismus festlegen. Wenn ein komplexes Mapping zu mehreren zusätzlichen Partitionspunkten führt, verwendet der Datenintegrationsdienst möglicherweise mehr Verarbeitungs-Threads, als die CPU verarbeiten kann.

Die Gesamtanzahl der Verarbeitungs-Threads ist gleich dem maximalen Parallelismuswert.

Beachten Sie die anderen Jobs, die der Datenintegrationsdienst ausführen muss.

Wenn Sie den maximalen Parallelismus so konfigurieren, dass jedes Mapping eine große Anzahl von Threads verwendet, sind für den Datenintegrationsdienst weniger Threads zur Ausführung zusätzlicher Jobs verfügbar.

Optional können Sie den Wert für ein Mapping ändern.

Standardmäßig ist der maximale Parallelismus für jedes Mapping auf „Auto“ gesetzt. Jedes Mapping verwendet den für den Datenintegrationsdienst definierten Wert für maximalen Parallelismus.

Im Developer Tool können Entwickler den Wert für den maximalen Parallelismus in den Laufzeiteigenschaften des Mappings ändern, um einen maximalen Wert für ein bestimmtes Mapping zu definieren. Wenn der maximale Parallelismus für den Datenintegrationsdienst und das Mapping auf zwei unterschiedliche Ganzzahlwerte festgelegt wurde, verwendet der Datenintegrationsdienst den Mindestwert.

Optimieren von Einfachdateien für Partitionierung

Wenn ein für die Partitionierung aktiviertes Mapping in einer Einfachdateiquelle liest oder in ein Einfachdateiziel schreibt, kann der Datenintegrationsdienst mehrere Threads verwenden, um in der Einfachdatei zu lesen oder in sie zu schreiben.

Optimieren von Einfachdateiquellen für die Partitionierung

Um eine optimale Leistung bei Verwendung mehrerer Threads zum Lesen aus einer Einfachdatei zu erzielen, konfigurieren Sie das Einfachdatei-Datenobjekt zur Optimierung des Durchsatzes, anstatt die Reihenfolge der Zeilen beizubehalten.

Zum Beheben von Engpässen bei partitionierten Einfachdateiquellen können Sie die folgende Lösung verwenden:

Konfigurieren Sie gleichzeitige Lesepartitionierung für das Einfachdatei-Datenobjekt zur Optimierung des Durchsatzes.

Legen Sie in den erweiterten Eigenschaften des Einfachdatei-Datenobjekts die Eigenschaft **Gleichzeitige Lesepartitionierung** fest, um den Durchsatz zu optimieren. Wenn Sie den Durchsatz optimieren, behält der Datenintegrationsdienst die Reihenfolge der Zeilen nicht bei, da die Zeilen in der Datei oder Dateiliste nicht der Reihe nach gelesen werden.

Optimieren von Einfachdateizielen für die Partitionierung

Um eine optimale Leistung bei der Verwendung mehrerer Threads zum Schreiben einer Einfachdatei zu erzielen, konfigurieren Sie Partitionen so, dass die Zielausgabe in separate Dateien geschrieben wird, und konfigurieren Sie mehrere Zielverzeichnisse.

Zum Beheben von Engpässen bei partitionierten Einfachdatei-Zielen können Sie die folgenden Lösungen verwenden:

Konfigurieren Sie Partitionen so, dass die Zielausgabe in separate Dateien geschrieben wird.

Legen Sie in den erweiterten Eigenschaften des Einfachdatei-Datenobjekts die Eigenschaft **Zusammenführungstyp** auf **Keine Zusammenführung** fest. Der Datenintegrationsdienst schreibt die Zielausgabe gleichzeitig in eine separate Datei für jede Partition. Wenn Sie zusammengeführte Zieldaten benötigen, optimiert der gleichzeitige Zusammenführungstyp die Leistung besser als der sequenzielle Zusammenführungstyp.

Konfigurieren Sie mehrere Zielverzeichnisse.

Wenn mehrere Threads in ein einzelnes Verzeichnis schreiben, kann das Mapping aufgrund eines E/A-Konflikts (Eingabe/Ausgabe) auf einen Engpass stoßen. Ein E/A-Konflikt kann auftreten, wenn Threads Daten gleichzeitig in das Dateisystem schreiben. Wenn Sie mehrere Verzeichnisse konfigurieren, legt der Datenintegrationsdienst das Ausgabeverzeichnis für jeden Thread im Ringverteilungsverfahren fest.

Konfigurieren Sie die Ausgabedateiverzeichnisse in den erweiterten Eigenschaften für das Einfachdatei-Datenobjekt. Verwenden Sie den TargetDir-Systemparameterwert, wenn ein Administrator mehrere durch Semikolons getrennte Verzeichnisse für die Eigenschaft **Zielverzeichnis** für den Datenintegrationsdienst im Administrator Tool eingegeben hat. Sie können auch einen anderen Wert zum Konfigurieren von mehreren Ausgabedateiverzeichnissen für das Einfachdatei-Datenobjekt eingeben.

Optimieren von relationalen Datenbanken für Partitionierung

Wenn ein für Partitionierung aktiviertes Mapping in einem IBM DB2 für LUW- oder in einem relationalen Oracle-Datenobjekt liest oder aber in dieses schreibt, kann der Datenintegrationsdienst mehrere Threads verwenden, um Daten in der relationalen Quelle zu lesen oder Daten in das relationale Ziel zu schreiben.

Um die Leistung bei der Verwendung mehrerer Threads, aus denen gelesen wird oder die in eine DB2 für LUW- oder eine relationale Datenbank geschrieben werden, zu verbessern, können Sie die Quell- und Zieltabellen partitionieren.

Hinweis: Wenn eine Zuordnung in einer relationalen Datenbank liest, bei der es sich nicht um DB2 für LUW oder Oracle handelt, oder in diese schreibt, verwendet der Datenintegrationsdienst einen Reader-Thread oder einen Writer-Thread.

Optimieren der Quelldatenbank für Partitionierung

Um eine optimale Leistung bei Verwendung mehrerer Threads zum Lesen aus einer DB2 für LUW oder einer Oracle-Quelldatenbank zu erzielen, vergewissern Sie sich, dass die Quelltabelle partitioniert und zum Akzeptieren paralleler Abfragen konfiguriert ist.

Um die Quelldatenbank für Partitionierung zu optimieren, führen Sie die folgenden Aufgaben durch:

Fügen Sie der Quelle Datenbankpartitionen hinzu.

Fügen Sie der relationalen Quelle Datenbankpartitionen hinzu, um die Geschwindigkeit der Datenintegrationsdienst-Abfrage zu erhöhen, die die Quelle liest. Wenn die Quelle keine Datenbankpartitionen aufweist, verwendet der Datenintegrationsdienst einen Thread zum Lesen aus der Quelle.

Aktivieren Sie paralleler Abfragen.

Relationale Datenbanken weisen möglicherweise Optionen auf, die parallele Abfragen für die Datenbank ermöglichen. Weitere Informationen zu diesen Optionen finden Sie in der Dokumentation zur Datenbank. Wenn diese Optionen nicht aktiviert sind, führt der Datenintegrationsdienst mehrere Partitions-SELECT-Anweisungen seriell aus.

Trennen Sie Daten in verschiedene Tablespaces.

Jede Datenbank enthält eine Option zum Trennen der Daten in verschiedene Tablespaces. Jeder Tablespace kann sich auf ein eindeutiges Dateisystem beziehen, wodurch E/A-Konflikte auf den Partitionen vermieden werden.

Erhöhen Sie die maximale Anzahl der in der Datenbank zulässigen Sitzungen.

Der Datenintegrationsdienst stellt eine separate Verbindung zur Quelldatenbank für jede Partition her. Erhöhen Sie die maximale Anzahl der zulässigen Sitzungen, damit die Datenbank eine größere Anzahl gleichzeitiger Verbindungen verarbeiten kann.

Optimieren der Zieldatenbank für Partitionierung

Um eine optimale Leistung bei Verwendung mehrerer Threads zum Schreiben in einer DB2 für LUW oder einer Oracle-Zieldatenbank zu erzielen, vergewissern Sie sich, dass die Zieltabelle partitioniert und zum parallelen Einfügen von Zeilen konfiguriert ist.

Um die Zieldatenbank für Partitionierung zu optimieren, führen Sie die folgenden Aufgaben durch:

Fügen Sie einem DB2 for LUW-Ziel Datenbankpartitionen hinzu.

Der Datenintegrationsdienst kann mehrere Threads zum Schreiben in ein DB2 für LUW-Ziel verwenden, das keine Datenbankpartitionen aufweist. Sie können die Ladeleistung jedoch optimieren, wenn das Ziel über Datenbankpartitionen verfügt. In diesem Fall stellt jeder Writer-Thread eine Verbindung zu dem DB2 für LUW-Knoten her, der die Datenbankpartition enthält. Da sich die Writer-Threads nicht alle mit einem einzelnen Master-Knoten verbinden, sondern eine Verbindung zu unterschiedlichen DB2 für LUW-Knoten herstellen, wird die Leistung erhöht.

Aktivieren Sie parallele Einfügungen.

Relationale Datenbanken weisen möglicherweise Optionen auf, die parallele Einfügungen in die Datenbank ermöglichen. Weitere Informationen zu diesen Optionen finden Sie in der Dokumentation zur Datenbank. Beispiel: Legen Sie die db_writer_processes-Option in einer Oracle-Datenbank und die max_agents-Option in einer DB2 für LUW-Datenbank fest, um parallele Einfügungen zu ermöglichen.

Trennen Sie Daten in verschiedene Tablespaces.

Jede Datenbank enthält eine Option zum Trennen der Daten in verschiedene Tablespaces. Jeder Tablespace kann sich auf ein eindeutiges Dateisystem beziehen, wodurch E/A-Konflikte auf den Partitionen vermieden werden.

Erhöhen Sie die maximale Anzahl der in der Datenbank zulässigen Sitzungen.

Der Datenintegrationsdienst stellt eine separate Verbindung zur Zieldatenbank für jede Partition her. Erhöhen Sie die maximale Anzahl der zulässigen Sitzungen, damit die Datenbank eine größere Anzahl gleichzeitiger Verbindungen verarbeiten kann.

Legen Sie Optionen zur Verbesserung der Datenbankskalierbarkeit fest.

Relationale Datenbanken weisen möglicherweise Optionen zur Verbesserung der Skalierbarkeit auf. Beispiel: Deaktivieren Sie die Archivprotokollierung und zeitlich vorgegebene Erstellung von Statistiken in einer Oracle-Datenbank, um die Skalierbarkeit zu verbessern.

Optimieren von Umwandlungen für Partitionierung

Wenn der Datenintegrationsdienst mehrere Threads zur Ausführung einer Aggregator-, Joiner-, Rang-, - oder Sortiererumwandlung einsetzt, verwendet der Dienst Cache-Partitionierung, um die Cachegröße zwischen den Threads aufzuteilen. Um die Leistung der Cache-Partitionierung zu optimieren, konfigurieren Sie mehrere Cache-Verzeichnisse.

Hinweis: Eine Lookup-Umwandlung kann nur ein einziges Cache-Verzeichnis verwenden.

Zum Beheben von Engpässen bei partitionierten Aggregator-, Joiner-, Rang- und Sortiererumwandlungen können Sie die folgende Lösung verwenden:

Konfigurieren Sie mehrere Cache-Verzeichnisse.

Bei der Cache-Partitionierung wird ein separater Cache für jede Partition erstellt, die eine Aggregator-, Joiner-, Rang-, - oder Sortiererumwandlung verarbeitet. Während der Cache-Partitionierung speichert jede Partition andere Daten in einem separaten Cache. Jeder Cache enthält die Zeilen, die von der jeweiligen Partition benötigt werden. Die Cache-Partitionierung optimiert die Mapping-Leistung, da jeder Thread parallel einen separaten Cache abfragt.

Wenn die Cachegröße kleiner als die für die Umwandlung erforderliche Speichergröße ist, schreiben die Umwandlungs-Threads Überlaufwerte in Cache-Dateien im Cache-Verzeichnis. Wenn mehrere Threads in ein einzelnes Verzeichnis schreiben, kann das Mapping aufgrund eines E/A-Konflikts auf einen Engpass

stoßen. Ein E/A-Konflikt kann auftreten, wenn Threads Daten gleichzeitig in das Dateisystem schreiben. Wenn Sie mehrere Cache-Verzeichnisse konfigurieren, legt der Datenintegrationsdienst das Ausgabeverzeichnis für jeden Umwandlungs-Thread im Ringverteilungsverfahren fest.

Konfigurieren Sie bei einer Aggregator-, Joiner- oder Rangumwandlung die Cache-Verzeichnisse in der erweiterten Eigenschaft **Cache-Verzeichnis**. Verwenden Sie den CacheDir-Systemparameterwert, wenn ein Administrator mehrere durch Semikolons getrennte Verzeichnisse für die Eigenschaft **Cache-Verzeichnis** für den Datenintegrationsdienst im Administrator Tool eingegeben hat. Sie können auch einen anderen Wert zum Konfigurieren von mehreren für die Umwandlung spezifischen Cache-Verzeichnissen eingeben.

Konfigurieren Sie bei einer Sortiererumwandlung die Cache-Verzeichnisse in der erweiterten Eigenschaft **Arbeitsverzeichnis**. Verwenden Sie den TempDir-Systemparameterwert, wenn ein Administrator mehrere durch Semikolons getrennte Verzeichnisse für die Eigenschaft **Temporäre Verzeichnisse** für den Datenintegrationsdienst im Administrator Tool eingegeben hat. Sie können auch einen anderen Wert zum Konfigurieren von mehreren für die Umwandlung spezifischen Cache-Verzeichnissen eingeben.

KAPITEL 7

Laufzeitoptimierung

Dieses Kapitel umfasst die folgenden Themen:

- Laufzeitoptimierung - Übersicht, 61
- Anwendungsdienstoptimierung, 61
- Überwachungsstatistiken, 63
- Speicherzuordnung, 65
- Datenobjekt-Caching, 66
- Systemoptimierung, 69

Laufzeitoptimierung - Übersicht

Aktivieren Sie Leistungsfunktionen und passen Sie Datenintegrationsdienst-Eigenschaften an, um die Leistung von Mappings zu optimieren.

Verwenden Sie die folgenden Optimierungsverfahren im Administrator Tool, um die besten Leistungsergebnisse für Ihre Anforderungen zu erzielen:

- Optimieren Sie Anwendungsdienstprozesse.
- Konfigurieren Sie Überwachungsstatistiken zum Überwachen von Systemengpässen.
- Reservieren Sie Speicher, um eine optimale Systemleistung zu erzielen.
- Konfigurieren des Datenobjekt-Caches
- Optimieren Sie das System, um Systemverzögerungen und langsamen Datenträgerzugriff zu vermeiden.

Anwendungsdienstoptimierung

Optimieren Sie den Anwendungsdienstprozess, wenn die Leistung beeinträchtigt ist. Sie können den Analyst-Dienst, Datenintegrationsdienst und den Modellrepository-Dienst optimieren.

Analyst Service-Optimierung

Passen Sie den Analyst Service an, um die Leistung zu optimieren. Sie können die Analyst Service-Prozesseigenschaft für den Arbeitsspeicher konfigurieren, die Netzwerk-Latenzzeit minimieren und die

Analyst Tool-Einstellungen für das Hochladen von Einfachdateien konfigurieren, um die Leistung des Service zu verbessern.

Zum Beheben von Analyst Service-Engpässen können Sie die folgenden Lösungen verwenden:

Legen Sie fest, dass das Analyst Tool zum Hochladen von Einfachdateien, die größer als 10 MB sind, eine Verbindung mit einem Speicherort im Netzwerk herstellt.

Die Leistung des Analyst-Diensts kann abnehmen, wenn Analysten Einfachdateien, die größer als 10 MB sind, in das Informatica-Installationsverzeichnis auf dem Computer hochladen, auf dem das Analyst-Tool ausgeführt wird. Dies kann sich auf Speicherplatz und Netzwerkleistung auswirken.

Laden Sie vom Analyst Tool Einfachdateien in das Informatica-Installationsverzeichnis hoch, die kleiner als 10 MB sind.

Die Leistung des Analyst Service-Prozesses kann sich verringern, wenn Analysten Einfachdateien, die größer als 10 MB sind, vom Analyst Tool in das Informatica-Installationsverzeichnis hochladen. Dies kann sich auf Speicherplatz und Netzwerkleistung auswirken.

Erhöhen Sie die Eigenschaft „Maximale Heap-Größe“ für den Analyst Service-Prozess.

Der Analyst Service-Prozess kann einen großen Betrag an Arbeitsspeicher verbrauchen, wenn eine große Anzahl von Benutzern gleichzeitig angemeldet ist. Dies kann dazu führen, dass eine große Anzahl von Netzwerkverbindungen zwischen dem Analyst Service und anderen Diensten, z. B. dem Data Integration Service oder dem Modell-Repository Service, geöffnet ist.

Legen Sie mit dem Administrator Tool in den erweiterten Eigenschaften für den Analyst Service-Prozess die Eigenschaft „Maximale Heap-Größe“ auf einen größeren Wert fest.

Exportieren Sie große Mapping-Spezifikationen in eine Tabelle oder in eine Einfachdatei und trunkieren Sie die Datei.

Die Leistung des Analyst Service-Prozesses kann beeinträchtigt werden, wenn Analysten große Mapping-Spezifikationen als Einfachdateien aus dem Analyst Tool exportieren.

Datenintegrationsdienst-Optimierung

Passen Sie den Datenintegrationsdienst an, um die Leistung des Dienstes zu verbessern. Sie können die Datenintegrationsdienst-Eigenschaften für den Arbeitsspeicher konfigurieren. Sie können jeden unter einem Datenintegrationsdienst ausgeführten Webdienst und SQL-Datendienst für die Behandlung von gleichzeitigen Anfragen konfigurieren.

Zum Beheben von Datenintegrationsdienst-Engpässen können Sie die folgenden Lösungen verwenden:

Konfigurieren Sie die Eigenschaft „Maximale Heap-Größe“ für den Datenintegrationsdienst-Prozess.

Der Datenintegrationsdienst kann einen großen Betrag an Arbeitsspeicher verbrauchen, während SQL-Datendienste und Webdienste verarbeitet werden.

Legen Sie mit dem Administrator Tool in den erweiterten Eigenschaften für den Datenintegrationsdienst-Prozess die Eigenschaft „Maximale Heap-Größe“ auf einen größeren Wert fest.

Konfigurieren Sie die Webdienst-Eigenschaft „DTM-Keep-Alive-Zeit“ für den Datenintegrationsdienst.

Der Datenintegrationsdienst verbraucht für jede Webdienst-Anfrage Systemressourcen zum Erzeugen einer DTM-Instanz. Legen Sie in der Konfiguration des Datenintegrationsdiensts fest, dass eine DTM-Instanz mehrere Webdienst-Anfragen verarbeitet.

Konfigurieren Sie mit dem Administrator Tool die Webdienst-Eigenschaft „DTM-Keep-Alive-Zeit“ für den Datenintegrationsdienst.

Konfigurieren Sie in den Eigenschaften des Datenintegrationsdienstes-Prozesses und den Eigenschaften des Webdienstes und des SQL-Datendienstes Ausführungsoptionen für gleichzeitige Anfragen.

Der Datenintegrationsdienst, jeder SQL-Datendienst und jeder Webdienst, der unter dem Datenintegrationsdienst ausgeführt wird, verbraucht für jede gleichzeitige Anfrage System- und Speicherressourcen.

Konfigurieren Sie die Eigenschaften des Datenintegrationsdienstes und des Webdiensts, um die Anzahl der gleichzeitigen Anfragen festzulegen, die der der Datenintegrationsdienst, jeder SQL-Datendienst und jeder Webdienst annehmen können.

Konfigurieren Sie mit dem Administrator Tool die folgenden Optionen und Eigenschaften für den Datenintegrationsdienst, Webdienst und SQL-Datendienst:

- Konfigurieren Sie die Ausführungsoptionen für den Datenintegrationsdienst.
- Konfigurieren Sie in den SQL-Eigenschaften die Eigenschaft „Maximale Anzahl an gleichzeitigen Verbindungen“ jedes SQL-Datendienstes für den Datenintegrationsdienst-Prozess.
- Konfigurieren Sie in den HTTP-Konfigurationseigenschaften für den Datenintegrationsdienst-Prozess die Eigenschaften „Maximale Anzahl an Backlog-Anfragen“ und „Maximale Anzahl an gleichzeitigen Anfragen“ für jeden Webdienst.

Deaktivieren Sie den Webdienst-Tracelevel.

Die Anzahl der Webdienst-Logdateien, die der Datenintegrationsdienst schreibt und verwaltet, kann die Leistung verringern.

Konfigurieren Sie mit dem Administrator Tool den Webdienst-Tracelevel, um die Anzahl der Webdienst-Laufzeit-Logdateien zu verringern, die der Datenintegrationsdienst auf Datenträger speichert.

Optimierung des Modell-Repository Service

Passen Sie den Analyst Service an, um die Leistung zu verbessern. Sie können die Modell-Repository Service-Prozesseigenschaft für den Arbeitsspeicher konfigurieren und die Netzwerk-Latenzzeit minimieren.

Zum Beheben von Modellrepository-Dienst-Engpässen können Sie die folgenden Lösungen verwenden:

Hosten Sie die Model Repository-Datenbank auf dem Computer, auf dem der Modell-Repository Service ausgeführt wird.

Die Modell-Repository Service-Prozessleistung kann beeinträchtigt werden, wenn die Model Repository-Datenbank auf einem Remote-Server gehostet wird. Modell-Repository Service-Vorgänge, die Kommunikation zwischen dem Model Repository und dem Modell-Repository Service in einem Netzwerk mit hoher Latenzzeit erfordern, können die Leistung des Modell-Repository Service beeinträchtigen.

Erhöhen Sie die Eigenschaft „Maximale Heap-Größe“ für den Modell-Repository Service-Prozess.

Der Modell-Repository Service-Prozess kann einen großen Betrag an Arbeitsspeicher verbrauchen, während eine große Anzahl von gleichzeitig angemeldeten Benutzern verarbeitet wird. Dies kann dazu führen, dass eine große Anzahl von Netzwerkverbindungen zwischen dem Modell-Repository Service und anderen Diensten, z. B. dem Data Integration Service oder dem Analyst Service, geöffnet ist.

Legen Sie mit dem Administrator Tool in den erweiterten Eigenschaften für den Modell-Repository Service-Prozess die Eigenschaft „Maximale Heap-Größe“ auf einen größeren Wert fest.

Überwachungsstatistiken

Bei der Überwachung handelt es sich um eine Domänenfunktion, die der Dienstmanager ausführt. Der Dienstmanager speichert die Überwachungskonfiguration im Modellrepository. Verwenden Sie im Administrator Tool die Registerkarte „Überwachen“ zum Überwachen von Systemengpässen, z. B. der Gesamtzahl von derzeit ausgeführten, fehlgeschlagenen, abgebrochenen und abgeschlossenen Jobs, die in einem ausgewählten Dienst ausgeführt werden.

Zum Beheben von Überwachungsstatistik-Engpässen können Sie die folgende Lösung verwenden:

Konfigurieren Sie die Domäne, um die Überwachung einzurichten.

Wenn Sie die Überwachung eingerichtet haben, speichert der Datenintegrationsdienst permanente Statistiken und Überwachungsberichte im Modellrepository. Permanente Statistiken sind historische Informationen über Integrationsobjekte, die zu einem früheren Zeitpunkt ausgeführt wurden. Die Überwachungsberichte zeigen Schlüsselmetriken zu einem Integrationsobjekt an.

Überwachungseinstellungen müssen Sie für die Domäne konfigurieren, um das Modellrepository anzugeben, in dem die Laufzeitstatistiken über die den Datenintegrationsdiensten zur Verfügung gestellten Objekte gespeichert sind. Die Überwachungseinstellungen gelten für alle Datenintegrationsdienste in der Domäne und können sich auf deren Leistung auswirken.

In der nachstehenden Tabelle sind Überwachungseinstellungen beschrieben, die sich auf die Leistung der Dienste auswirken können:

Option	Beschreibung
Anzahl an Tagen für die Aufbewahrung von historischen Daten	Anzahl an Tagen von historischen Statistiken, die im Modellrepository gespeichert sind.
Statistik bereinigen alle	Häufigkeit in Tagen, mit der der Datenintegrationsdienst die Statistiken bereinigt. Der Standardwert ist 1.
Tage bei	Tageszeit, zu der der Datenintegrationsdienst alte Statistiken bereinigt. Der Standardwert ist 1:00 morgens.
Maximale Anzahl an sortierbaren Datensätzen	Maximale Anzahl an Datensätzen, die Sie auf der Registerkarte **Überwachen** sortieren können. Übersteigt die Anzahl der Datensätze auf der Registerkarte **Überwachen** diesen Wert, können Sie nur die Spalten **Startzeit** und **Endezeit** sortieren. Der Standardwert ist 3.000.
Maximale Verzögerung für Aktualisierungsbenachrichtigungen	Maximaler Zeitraum in Sekunden, für den der Datenintegrationsdienst die Statistiken zwischenspeichert, bevor er sie dauerhaft im Modellrepository ablegt und auf der Registerkarte **Überwachen** anzeigt. Standardwert ist 10.
Millisekunden einblenden	Auf der Registerkarte **Überwachen** werden bei Datums- und Zeitfeldern auch Millisekunden angezeigt.

Speicherzuordnung

Um die Mapping-Leistung zu optimieren, konfigurieren Sie Speichereigenschaften für den Datenintegrationsdienst im Administrator Tool.

Die folgende Tabelle beschreibt die Eigenschaft „Maximale Speichergröße pro Anfrage“ für das Zuordnungsdienstmodul:

Eigenschaft	Beschreibung
Maximale Speichergröße pro Anfrage	Das Verhalten von „Maximale Speichergröße pro Anfrage“ richtet sich nach den folgenden Datenintegrationsdienst-Konfigurationen: - Der Dienst führt Jobs in lokalen Prozessen oder Remoteprozessen aus oder die Diensteigenschaft „Maximale Speichergröße“ lautet 0 (Standardeinstellung). „Maximale Speichergröße pro Anfrage“ ist die maximale Speichermenge in Byte, die der Datenintegrationsdienst allen Umwandlungen zuordnen kann, die den automatischen Cache-Modus in einer einzelnen Anfrage verwenden. Der Dienst weist Arbeitsspeicher separat zu Umwandlungen zu, die über eine bestimmte Cache-Größe verfügen. Der von der Anfrage verwendete Gesamtspeicher kann den Wert für „Maximale Speichergröße pro Anfrage“ überschreiten. - Der Dienst führt Jobs in dem Prozess des Datenintegrationsdiensts aus und die Diensteigenschaft „Maximale Speichergröße“ ist größer als 0. „Maximale Speichergröße pro Anfrage“ ist die maximale Speichermenge in Byte, die der Datenintegrationsdienst einer einzelnen Anfrage zuordnen kann. Der von der Anfrage verwendete Gesamtspeicher darf den Wert für „Maximale Speichergröße pro Anfrage“ nicht überschreiten. Standardwert ist 536.870.912.

In der folgenden Tabelle werden die Ausführungsoptionen für den Datenintegrationsdienst beschrieben:

Eigenschaft	Beschreibung
Maximale Ausführungspoolgröße	Maximale Anzahl an Jobs, die pro Datenintegrationsdienstprozess gleichzeitig ausgeführt werden können. Zu den Jobs gehören Datenvorschauen, Zuordnungen, Profilerstellungsjobs, SQL-Abfragen und Webdienstanfragen. Ein Datenintegrationsdienstgitter umfasst beispielsweise drei laufende Dienstprozesse. Wenn Sie den Wert auf 10 festlegen, kann jeder Datenintegrationsdienst bis zu 10 Jobs gleichzeitig ausführen. Insgesamt 30 Jobs können gleichzeitig auf dem Gitter ausgeführt werden. Standardwert ist 10. **Hinweis:** Wenn Sie den Wert für die Poolgröße erhöhen, verwendet der Datenintegrationsdienst mehr Hardware-Ressourcen wie CPU, Arbeitsspeicher und System-E/A. Legen Sie diesen Wert basierend auf den auf dem Computer verfügbaren Ressourcen fest. Berücksichtigen Sie beispielsweise die Anzahl der CPUs auf dem Computer, der den Datenintegrationsdienst hostet, und die Menge an Arbeitsspeicher, die für den Datenintegrationsdienst verfügbar ist.
Maximale Speichergröße	Die maximale Speichermenge in Byte, die der Datenintegrationsdienst für die gleichzeitige Ausführung aller Anfragen zuordnen kann, wenn der Dienst Jobs in dem Prozess des Datenintegrationsdiensts ausführt. Wenn der Datenintegrationsdienst Jobs in separaten lokalen Prozessen oder Remoteprozessen ausführt, ignoriert der Dienst diesen Wert. Wenn Sie die Speichergröße, die der Datenintegrationsdienst zuordnen kann, nicht einschränken möchten, legen Sie diese Eigenschaft auf 0 fest. Wenn der Wert größer als 0 ist, verwendet der Datenintegrationsdienst die Eigenschaft zur Berechnung des maximalen Gesamtspeicherplatzes, der für die gleichzeitige Ausführung aller Anfragen zulässig ist. Der Datenintegrationsdienst berechnet den maximalen Gesamtspeicherplatz folgendermaßen: Maximale Speichergröße + maximale Heap-Größe + zum Laden von Programmkomponenten erforderlicher Speicherplatz Standardwert ist 0. **Hinweis:** Wenn Sie Profile oder Datenqualitäts-Mappings ausführen, müssen Sie diese Eigenschaft auf 0 festlegen.

Die folgende Tabelle beschreibt die Eigenschaft „Maximale Heap-Größe“ für den Datenintegrationsdienst-Prozess:

Eigenschaft	Beschreibung
Maximale Heap-Größe	RAM-Größe für die Java Virtual Machine (JVM), auf der der Datenintegrationsdienst ausgeführt wird Mit dieser Eigenschaft verbessern Sie die Leistung. Fügen Sie einen der folgenden Buchstaben an den Wert an, um die Einheiten anzugeben: - b für Byte. - k für Kilobyte - m für Megabyte - g for gigabytes Standardwert ist 640 Megabyte. **Hinweis:** Sie können die maximale Heap-Größe erhöhen, wenn der Datenintegrationsdienst große Mengen von Daten verarbeiten muss.

Datenobjekt-Caching

Der Datenintegrationsdienst verwendet Datenobjekt-Caching für den Zugriff auf vorgefertigte logische Datenobjekte und virtuelle Tabellen. Aktivieren Sie Datenobjekt-Caching, um die Leistung für Mappings, SQL-

Datendienstabfragen und Webdienstanfragen zu erhöhen, die logische Datenobjekte und virtuelle Tabellen enthalten.

Standardmäßig extrahiert der Datenintegrationsdienst Quelldaten und erstellt benötigte Datenobjekte, wenn er ein Mapping, eine SQL-Datendienstabfrage oder eine Webdienstanfrage ausführt. Wenn Sie Datenobjekt-Caching aktivieren, kann der Datenintegrationsdienst auf logische Datenobjekte und virtuelle Tabellen im Cache zugreifen.

Führen Sie die folgenden Schritte aus, um Datenobjekt-Caching für logische Datenobjekte und virtuelle Tabellen in einer Anwendung zu konfigurieren:

1. Konfigurieren Sie die Datenbankverbindung für den Datenobjekt-Cache in den Cache-Eigenschaften für den Datenintegrationsdienst.
2. Aktivieren Sie Caching in den Eigenschaften der logischen Datenobjekte oder virtuellen Tabellen in einer Anwendung.

Die Datenobjekt-Cache-Manager-Komponente des Datenintegrationsdiensts verwaltet die Cache-Tabellen für logische Datenobjekte und virtuelle Tabellen in der Datenobjekt-Cache-Datenbank. Wenn der Datenobjekt-Cache-Manager den Cache verwaltet, fügt er bei jeder Aktualisierung alle Daten in die Cache-Tabellen ein. Wenn Sie die Cache-Tabellen schrittweise aktualisieren möchten, können Sie die Cache-Tabellen mit einem Datenbank-Client oder einem anderen externen Tool selbst verwalten. Nach dem Aktivieren des Datenobjekt-Caching können Sie ein logisches Datenobjekt oder eine virtuelle Tabelle zur Verwendung einer benutzerverwalteten Cache-Tabelle konfigurieren.

Um den Datentyp „Zeitstempel mit Zeitzone“ zu verwenden und das Datenobjekt-Caching für IBM DB2 oder Microsoft SQL Server zu aktivieren, setzen Sie das Datums- und Zeitformat des bereitgestellten Mappings auf das Format „YYYY-MM-DD HH24:MI:SS“. Der Datenintegrationsdienst schreibt die Daten sekundengenau.

Datentypen für Cache-Tabellen

Der Datenintegrationsdienst verwendet Daten aus Cache-Tabellen, wenn er Mappings, SQL-Datendienstabfragen und Webdienstanfragen verarbeitet, die zwischengespeicherte Objekte enthalten. Die Cache-Tabellen-Datentypen, die der Datenintegrationsdienst erwartet, können von den zwischengespeicherten Objekt-Datentypen abweichen.

Der Datenobjekt-Cache-Manager erstellt die Cache-Tabellen mit den Datentypen, die der Datenintegrationsdienst erwartet. Wenn Sie benutzerverwaltete Cache-Tabellen verwenden, stellen Sie sicher, dass die Cache-Tabellen die Datentypen verwenden, die der Datenintegrationsdienst erwartet.

Cache-Datentypen für virtuelle Tabellen

In der folgenden Tabelle finden Sie eine Auflistung der Cache-Tabellendatentypen für virtuelle Tabellen:

Datentyp für virtuelle Tabellen	IBM DB2	Microsoft SQL Server	Oracle
Char	Vargraphic Dbclob, für Präzision größer als 32672	Nvarchar Ntext, für Präzision größer als 4000	Nvarchar2 Nclob, für Präzision größer als 2000
Bigint	Bigint	Bigint	Zahl
Boolescher Wert	Integer	Int	Zahl
Datum	Zeitstempel	Datetime2	Zeitstempel

Datentyp für virtuelle Tabellen	IBM DB2	Microsoft SQL Server	Oracle
Double	Double	Fließkomma	Zeitstempel
Dezimal	Dezimal	Dezimal	Zahl
Int	Integer	Int	Zahl
Uhrzeit	Zeitstempel	Datetime2	Zeitstempel
Zeitstempel	Zeitstempel	Datetime2	Zeitstempel
Varbinary	Blob	Binär Bild, für Präzision größer als 8000	Unformatiert Blob, für Präzision größer als 2000
Varchar	Vargraphic Dbclob, für Präzision größer als 32672	Nvarchar Ntext, für Präzision größer als 4000	Nvarchar2 Nclob, für Präzision größer als 2000

Cache-Datentypen für logische Datenobjekte

In der folgenden Tabelle finden Sie eine Auflistung der Cache-Tabellendatentypen für logische Datenobjekte:

Datentyp für logische Datenobjekte	DB2	Microsoft SQL Server	Oracle
Bigint	Bigint	Bigint	Zahl
Binär	Blob	Binär Bild, für Präzision größer als 8000	Unformatiert Blob, für Präzision größer als 2000
Datum/Uhrzeit	Zeitstempel	Datetime2	Zeitstempel
Double	Double	Fließkomma	Zahl
Dezimal	Dezimal	Dezimal	Zahl
Integer	Integer	Int	Zahl
String	Vargraphic Dbclob, für Präzision größer als 32672	Nvarchar Ntext, für Präzision größer als 4000	Nvarchar2 Nclob, für Präzision größer als 2000
Text	Vargraphic Dbclob, für Präzision größer als 32672	Nvarchar Ntext, für Präzision größer als 4000	Nvarchar2 Nclob, für Präzision größer als 2000

Cache-Optimierung für Datenobjekte

Die Cache-Leistung richtet sich nach der Leistung der Cache-Datenbank und der Konfiguration der Objekte innerhalb von Mappings, SQL-Datendiensten und Webdiensten.

Berücksichtigen Sie die folgenden Lösungen, um die Cache-Leistung zu erhöhen:

Optimieren Sie die Cache-Datenbank.

Die optimale Leistung für den Cache hängt von der Geschwindigkeit und der Leistung der Cache-Datenbank sowie der Cachegröße ab. Konfigurieren Sie die Cachegröße innerhalb der Cache-Datenbank.

Da der Data Object Cache Manager den alten Cache für einen Aktualisierungsvorgang beibehalten muss, muss die Größe des Caches zum Speichern von zwei Datensätzen ausreichen. Mit der folgenden Formel können Sie einen Näherungswert für die erforderliche Mindestgröße des Caches berechnen:

```
2 * average data object size * number of data objects
```

Beispiel: Sie möchten 20 logische Datenobjekte und 10 virtuelle Tabellen im Cache speichern. Wenn die durchschnittliche Objektgröße 15 MB beträgt, ist die erforderliche Cachegröße 2 * 15 MB * (20 + 10) = 900 MB.

Cache-Tabellen sind schreibgeschützt. Endbenutzer können die Cache-Tabellen nicht mit SQL-Befehlen aktualisieren.

Definieren Sie Primärschlüssel und Fremdschlüssel für logische Datenobjekte.

Wenn der Datenintegrationsdienst den Cache für logische Datenobjekte generiert, erstellt er Indizes. Die Indizes können die Leistung von Anfragen bei der Cache-Datenbank erhöhen.

Zwischenspeichern Sie logische Datenobjekte, die Sie in einem Mapping verknüpfen.

Wenn Sie zwischengespeicherte logische Datenobjekte verknüpfen, kann der Datenintegrationsdienst die Joiner-Umwandlungslogik in die Cache-Datenbank übertragen, auch wenn die Quelldaten aus unterschiedlichen Datenbanken stammen.

Generieren Sie den Index-Cache basierend auf Spalten in einem logischen Datenobjekt oder einer virtuelle Tabelle.

Konfigurieren Sie den Datenintegrationsdienst, um einen Index-Cache basierend auf Spalten in einem logischen Datenobjekt oder einer virtuellen Tabelle zu generieren. Der Index kann die Leistung von Abfragen in der Cache-Datenbank erhöhen.

Systemoptimierung

Häufig kommt es zu einer Leistungsverringerung, weil das Mapping ineffiziente Verbindungen oder ein überlastetes Datenintegrationsdienst-Prozesssystem verwendet. Systemverzögerungen können auch durch Router, Switches, Netzwerkprotokolle und die Nutzung durch viele Benutzer verursacht werden.

Die Mapping-Leistung kann durch langsamen Datenträgerzugriff auf Quell- und Target-Datenbanken, Quell- und Target-Dateisysteme und Knoten in der Domäne verursacht werden. Lassen Sie Systemadministrator die Festplatten auf den Computern überprüfen.

Zum Beheben von Systemoptimierungsengpässen können Sie die folgenden Lösungen verwenden:

Erhöhen Sie die Netzwerkgeschwindigkeit.

Langsame Netzwerkverbindungen können die Mapping-Leistung verringern. Lassen Sie den Systemadministrator bestimmen, ob das Netzwerk mit optimaler Geschwindigkeit ausgeführt wird. Verringern Sie die Anzahl der Netzwerk-Hops zwischen dem Datenintegrationsdienst-Prozess und den Datenintegrationsdienst-Datenbanken.

Verwenden Sie mehrere CPUs.

Sie können mehrere CPUs verwenden, um mehrere Mappings gleichzeitig auszuführen.

Reduzieren Sie Paging.

Wenn der physische Arbeitsspeicher eines Betriebssystems nicht mehr ausreicht, wird Speicher auf den Datenträger ausgelagert, um physischen Arbeitsspeicher freizugeben. Konfigurieren Sie den physischen Arbeitsspeicher für den Datenintegrationsdienst-Prozess, um die Auslagerung auf den Datenträger zu minimieren.

Verwenden Sie Prozessorbindung.

In einer UNIX-Umgebung mit mehreren Prozessoren verwendet der Datenintegrationsdienst möglicherweise eine große Menge von Systemressourcen. Steuern Sie mit Hilfe von Prozessorbindung die Prozessornutzung durch den Datenintegrationsdienst-Prozess. Wenn sich Quell- und Target-Datenbank auf demselben Computer befinden, beschränken Sie außerdem mit Hilfe von Prozessorbindung die von der Datenbank verwendeten Ressourcen.

KAPITEL 8

SQL-Datendienstoptimierung

Dieses Kapitel umfasst die folgenden Themen:

- SQL-Datendienstoptimierung - Übersicht, 71
- Optimieren von Fremdanbieter-Client-Tools, 72
- Optimierungsebenen des SQL-Datendiensts, 72
- Eigenschaften des SQL-Datendiensts für Arbeitsspeicher und gleichzeitige Anfragen, 76
- Ergebnissatz-Cache für einen SQL-Datendienst, 77
- Dauerhaft virtuelle Daten in temporären Tabellen, 78

SQL-Datendienstoptimierung - Übersicht

Sie können SQL-Datendienste optimieren, um die Leistung zu verbessern, wenn Endbenutzer mit Fremdanbieter-Client-Tools SQL-Abfragen dieser Datendienste ausführen. Wenn ein SQL-Datendienst ein virtuelles Tabellen-Mapping verwendet, können Sie die Quellen, die Umwandlungen und das Mapping optimieren.

Verwenden Sie zum Optimieren eines SQL-Datendienstes die folgenden Optimierungsverfahren:

- Optimieren von Fremdanbieter-Client-Tools
- Konfigurieren Sie die SQL-Datendienst-Optimierungsebene.
- Konfigurieren Sie SQL-Datendiensteigenschaften für Parallelität und Speicher eines Datendienstprozesses.
- Konfigurieren Sie die Datenobjekt-Zwischenspeicherung für den SQL-Datendienst.
- Konfigurieren Sie das Ergebnis-Caching für den SQL-Datendienst.
- Konfigurieren Sie Einschränkungen für virtuelle Tabellen im SQL-Datendienst.

Verwandte Themen:

- "Datenobjekt-Caching" auf Seite 66

Optimieren von Fremdanbieter-Client-Tools

Fremdanbieter-Client-Tools können die Leistung beeinträchtigen, wenn SQL-Abfragen für einen SQL-Datendienst verarbeitet und ausgeführt werden. Fremdanbieter-Client-Tools können die Leistung beeinträchtigen, wenn SQL-Abfragen für einen SQL-Datendienst verarbeitet und ausgeführt werden.

Zum Beheben von Engpässen bei Fremdanbieter-Client-Tools können Sie die folgenden Lösungen verwenden:

Senden Sie umfangreiche Abfrageergebnisse in eine Datei auf dem Datenträger.

Ein Fremdanbieter-Client-Tool kann die Leistung beeinträchtigen, wenn es umfangreiche Abfrageergebnisse im Konsolenfenster anzeigt.

Deaktivieren Sie im Fremdanbieter-Client-Tool die Verschlüsselung.

Ein Fremdanbieter-Client-Tool kann die Leistung beeinträchtigen, wenn es während des Abrufs oder Anzeigens von Abfrageergebnissen Daten verschlüsselt.

Legen Sie fest, dass das Fremdanbieter-Client-Tool vorab einen Satz von Zeilen abruft.

Ein Fremdanbieter-Client-Tool kann die Leistung beeinträchtigen, wenn es Zeilen einzeln nacheinander abruft.

Deaktivieren Sie im Fremdanbieter-Client-Tool die Option zum Lesen des Inhalts der Tabelle beim ersten Laden der Tabelle.

Ein Fremdanbieter-Client-Tool kann die Leistung beeinträchtigen, wenn für den BLOB- und CLOB-Datentyp das Lesen des Inhalts der Tabelle beim ersten Laden der Tabelle festgelegt ist und diese Datentypen nicht in der Abfrage verwendet werden.

Legen Sie im Fremdanbieter-Client-Tool die Standardformat- und Standardkonvertierungseinstellungen für Datum, Uhrzeit und Zeitstempel fest.

Ein Fremdanbieter-Client-Tool kann die Leistung beeinträchtigen, wenn statt der Standardeinstellungen benutzerdefinierte Format- und Konvertierungseinstellungen festgelegt sind.

Deaktivieren Sie die Debug-Option (bzw. legen Sie fest, dass Debugging nicht ausgeführt wird).

Ein Fremdanbieter-Client-Tool kann die Leistung beeinträchtigen, wenn die Debug-Option für die Abfrage auf Einzelschritt festgelegt ist. Dies kann die Leistung verringern, da das Fremdanbieter-Client-Tool während des Verarbeitens der Abfrage mehr Log-Meldungen in die Debug-Datei schreibt.

Optimierungsebenen des SQL-Datendiensts

Der Data Integration Service optimiert SQL-Datendienste, die auf der Optimierungsebene basieren, die Sie konfigurieren. Konfigurieren Sie die Optimierungsebene, wenn Sie möchten, dass der SQL-Datendienst eine Optimierungsebene verwendet, die von der normalen Optimierungsebene abweicht. Standardmäßig verwendet jeder SQL-Datendienst die normale Optimierungsebene.

Weitere Informationen darüber, wie die Optimierungsebene eine optimierte Abfrage für einen SQL-Datenservice erstellt, finden Sie im Abfrageplan für einen SQL-Datendienst. Wenn Sie den Abfrageplan sehen,

zeigt das Developer-Tool eine grafische Darstellung der optimierten Abfrage basierend auf der Optimierungsebene und einer grafischen Darstellung der ursprünglichen Abfrage an.

Sie können die folgenden Optimierungsebenen konfigurieren:

Keiner

Der Datenintegrationsdienst wendet keine Optimierung an.

Minimal

Der Datenintegrationsdienst wendet die frühe Projektionsoptimierungsmethode an.

Normal

Der Datenintegrationsdienst wendet folgende Optimierungsmethoden an: frühe Projektion, frühe Auswahl, Verzweigungsbereinigung, Push-Into, globale Vorhersage und Vorhersage. Normal ist die Standard-Optimierungsebene.

Komplett

Der Datenintegrationsdienst wendet folgende Optimierungsmethoden an: kostenbasiert, frühe Projektion, frühe Auswahl, Verzweigungsbereinigung, Vorhersage, Push-Into, Semi-Join und Dataship-Join.

Sie können eine oder mehrere der folgenden Methoden zum Konfigurieren der Optimierungsebene für einen SQL-Datendienst verwenden:

- Konfigurieren Sie die Optimierungsebene für die Datenvorschau der SQL-Datendienste.
- Konfigurieren Sie die Optimierungsebene für bereitgestellte SQL-Datendienste.
- Konfigurieren Sie die Optimierungsebene in den Verbindungszeichenfolgen von Abfragen, die Sie gegen einen bereitgestellten SQL-Datendienst ausführen.

Konfigurieren der SQL-Datendienst-Optimierungsebene für die Datenvorschau

Konfigurieren Sie die Optimierungsebene, die der Data Integration Service verwendet, um die SQL-Abfragen auszuführen, wenn Sie die Vorschau der Ausgabe eines SQL-Datendienstes sehen.

1. Klicken Sie im Developer-Tool auf **Ausführen** > **Öffnen des ausgeführten Dialogs**.

   Das Dialogfeld **Ausführen** wird angezeigt.
2. Klicken Sie auf **Daten-Viewer-Konfiguration**.
3. Klicken Sie auf die Schaltfläche **Neu**.
4. Geben Sie einen Namen für die Daten-Viewer-Konfiguration ein.
5. Klicken Sie auf die Registerkarte **Erweitert**.
6. Wählen Sie eine Optimierungsebene.
7. Klicken Sie auf **Anwenden**.
8. Klicken Sie auf **Schließen**

   Das Developer-Tool erstellt die Daten-Viewer-Konfiguration.

Konfigurieren der Optimierungsebene für bereitgestellte SQL Data Services

Konfigurieren Sie die Optimierungsebene, die die Data Integration Services verwenden, um die SQL-Abfragen gegen einen bereitgestellten SQL-Datendienst auszuführen. Sie können auswählen, die Optimierungsebene

für eine einzelne Abfrage zu überschreiben, indem Sie die Optimierungsebene in der SQL-Datendienst-Verbindung konfigurieren.

1. Wählen Sie im Administratortool einen Data Integration Service aus.
2. Klicken Sie auf die Ansicht **Anwendungen**.
3. Erweitern Sie die Anwendung, die den SQL-Datendienst enthält, für den Sie die Optimierungsebene konfigurieren möchten.
4. Wählen Sie den SQL-Datendienst aus und bearbeiten Sie die folgende Eigenschaft:

Eigenschaft	Beschreibung
Optimierungslevel	Die Optimierungsebene, die der Data Integration Service für das Objekt anwendet. Geben Sie den numerischen Wert ein, der mit der Optimierungsebene verbunden ist, die Sie konfigurieren möchten. Sie können Sie einen der folgenden numerischen Werte eingeben: - 0. Der Datenintegrationsdienst wendet keine Optimierung an. - 1. Der Datenintegrationsdienst wendet die frühe Projektionsoptimierungsmethode an. - 2. Der Datenintegrationsdienst wendet die Optimierungsmethoden „Frühe Projektion“, „Frühe Auswahl“, „Push-Into“ und „Prädikat“ an. - 3. Der Datenintegrationsdienst wendet die Optimierungsmethoden „Kostenbasiert“, „Frühe Projektion“, „Frühe Auswahl“, „Push-Into“, „Prädikat“ und „Semi-Join“ an.

5. Um die Optimierungsebene zu überschreiben, die der Data Integration Services zum Ausführen einer Abfrage verwendet, fügen Sie den folgenden Eintrag in die JDBC-URL- oder ODBC-Verbindungszeichenfolge ein: `SQLDataServiceOptions.optimizeLevel= <numeric_optimizer_level>`.

SQL-Datendienstabfrageplan

Wenn Sie den Abfrageplan für einen SQL-Datendienst einsehen, sehen Sie die grafische Darstellung der ursprünglichen Abfrage und die grafische Darstellung der optimierten Abfrage. Die grafische Darstellung beschreibt, wie der Data Integration Service die Abfrage verarbeitet. Sie enthält die Umwandlungen und die Reihenfolge, in der der Data Integration Service jede Umwandlung verarbeitet.

Das Developer-Tool verwendet die Optimierungsebene, die Sie im Developer-Tool zum Generieren der optimierten Abfrage festgelegt haben. Die optimierte Abfrage zeigt die Abfrage an, während der Data Integration Service sie ausführt.

Sie möchten beispielsweise die virtuelle Tabelle des KUNDEN in einem SQL-Datendienst abfragen. In der Ansicht **Daten-Viewer** wählen Sie die Standardkonfigurationseinstellungen für den Daten-Viewer aus, welche die Optimierungsebene für die Abfrage auf „Normal“ festlegt.

Geben Sie folgende Abfrage in die Ansicht **Daten-Viewer** ein:

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

Wenn Sie den SQL-Abfrageplan anzeigen, zeigt das Developer-Tool die folgende grafische Darstellung der Abfrage an:

Die nicht optimierte Ansicht zeigt die Abfrage an, die Sie eingeben. Das Developer-Tool zeigt die WHERE-Klausel als Filterumwandlung und die ORDER BY-Klausel als Sortierumwandlung an. Das Developer-Tool verwendet die pass-through Ausdrucksumwandlung zur Umbenennung von Ports.

Wenn Sie die optimierte Abfrage sehen, zeigt das Developer-Tool die folgende grafische Darstellung der Abfrage an:

Die optimierte Ansicht zeigt die Abfrage, die der Data Integration Service ausführt. Da die Optimierungsebene normal ist, verschiebt der Data Integration Service die Filterbedingung zum Datenquellobjekt. Das Verschieben der Filterbedingung verbessert die Abfrageleistung, da die Anzahl der Zeilen, die der Data Integration Service vom Quelldatenobjekt liest, reduziert wird. Wie in der nicht optimierten Abfrage, zeigt das Developer-Tool die ORDER BY-Klausel als Sortierumwandlung an. Es werden pass-through-Ausdrucksumwandlungen verwendet, um die Datentypen dazu zu zwingen, die logischen Umwandlungen anzugeben.

SQL-Abfrageplan anzeigen

Lassen Sie einen SQL-Abfrageplan anzeigen, um eine zuordnungsähnliche Repräsentation der erstellen SQL-Abfrage zu erhalten.

1. Öffnen Sie einen SQL-Datendienst, der mindestens eine virtuelle Tabelle enthält.
2. Klicken Sie auf die Ansicht **Daten-Viewer**.
3. Geben Sie eine SQL-Abfrage in das Fenster **Eingabe** ein.
4. Sie können eine Daten-Viewer-Konfiguration wählen, die jene Optimierungsebene enthält, die Sie für die Abfrage verwenden möchten.
5. Klicken Sie auf **Abfrageplan zeigen**.

   Das Developer Tool zeigt den SQL-Abfrageplan für die Abfrage so an, wie Sie dies in der Registerkarte **Nicht-Optimiert** eingegeben haben.
6. Um die optimierte Abfrage anzuzeigen, klicken Sie auf die Registerkarte **Optimiert**.

   Das Developer Tool zeigt den optimierten SQL-Abfrageplan an.

Eigenschaften des SQL-Datendiensts für Arbeitsspeicher und gleichzeitige Anfragen

Um die Leistung des SQL-Datendiensts zu optimieren, konfigurieren Sie im Administrator Tool Parallelitäts- und Arbeitsspeicher-Eigenschaften für den Datenintegrationsdienst.

Die folgende Tabelle beschreibt die Eigenschaft „Maximale Speichergröße pro Anfrage“ für das SQL-Dienstmodul:

Eigenschaft	Beschreibung
Maximale Speichergröße pro Anfrage	Das Verhalten von „Maximale Speichergröße pro Anfrage“ richtet sich nach den folgenden Datenintegrationsdienst-Konfigurationen: - Der Dienst führt Jobs in lokalen Prozessen oder Remoteprozessen aus oder die Diensteigenschaft „Maximale Speichergröße“ lautet 0 (Standardeinstellung). „Maximale Speichergröße pro Anfrage“ ist die maximale Speichermenge in Byte, die der Datenintegrationsdienst allen Umwandlungen zuordnen kann, die den automatischen Cache-Modus in einer einzelnen Anfrage verwenden. Der Dienst weist Arbeitsspeicher separat zu Umwandlungen zu, die über eine bestimmte Cache-Größe verfügen. Der von der Anfrage verwendete Gesamtspeicher kann den Wert für „Maximale Speichergröße pro Anfrage“ überschreiten. - Der Dienst führt Jobs in dem Prozess des Datenintegrationsdiensts aus und die Diensteigenschaft „Maximale Speichergröße“ ist größer als 0. „Maximale Speichergröße pro Anfrage“ ist die maximale Speichermenge in Byte, die der Datenintegrationsdienst einer einzelnen Anfrage zuordnen kann. Der von der Anfrage verwendete Gesamtspeicher darf den Wert für „Maximale Speichergröße pro Anfrage“ nicht überschreiten. Standardwert ist 50.000.000.

Die folgende Tabelle beschreibt die Eigenschaft „Maximale Heap-Größe“ für den Datenintegrationsdienst-Prozess:

Eigenschaft	Beschreibung
Maximale Heap-Größe	RAM-Größe für die Java Virtual Machine (JVM), auf der der Datenintegrationsdienst ausgeführt wird Mit dieser Eigenschaft verbessern Sie die Leistung. Fügen Sie einen der folgenden Buchstaben an den Wert an, um die Einheiten anzugeben: - b für Byte. - k für Kilobyte - m für Megabyte - g for gigabytes Standardwert ist 640 Megabyte. **Hinweis:** Sie können die maximale Heap-Größe erhöhen, wenn der Datenintegrationsdienst große Mengen von Daten verarbeiten muss.

Die folgende Tabelle beschreibt die SQL-Eigenschaften für den Datenintegrationsdienstprozess:

Eigenschaft	Beschreibung
Maximale Anzahl an gleichzeitigen Verbindungen	Begrenzt die Anzahl der Datenbankverbindungen, die der Datenintegrationsdienst für SQL-Datendienste herstellen kann. Standardwert ist 100.

In der folgenden Tabelle werden die Ausführungsoptionen für den Datenintegrationsdienst beschrieben:

Eigenschaft	Beschreibung
Maximale Ausführungspoolgröße	Maximale Anzahl an Jobs, die pro Datenintegrationsdienstprozess gleichzeitig ausgeführt werden können. Zu den Jobs gehören Datenvorschauen, Zuordnungen, Profilerstellungsjobs, SQL-Abfragen und Webdienstanfragen. Ein Datenintegrationsdienstgitter umfasst beispielsweise drei laufende Dienstprozesse. Wenn Sie den Wert auf 10 festlegen, kann jeder Datenintegrationsdienst bis zu 10 Jobs gleichzeitig ausführen. Insgesamt 30 Jobs können gleichzeitig auf dem Gitter ausgeführt werden. Standardwert ist 10. **Hinweis:** Wenn Sie den Wert für die Poolgröße erhöhen, verwendet der Datenintegrationsdienst mehr Hardware-Ressourcen wie CPU, Arbeitsspeicher und System-E/A. Legen Sie diesen Wert basierend auf den auf dem Computer verfügbaren Ressourcen fest. Berücksichtigen Sie beispielsweise die Anzahl der CPUs auf dem Computer, der den Datenintegrationsdienst hostet, und die Menge an Arbeitsspeicher, die für den Datenintegrationsdienst verfügbar ist.
Maximale Speichergröße	Die maximale Speichermenge in Byte, die der Datenintegrationsdienst für die gleichzeitige Ausführung aller Anfragen zuordnen kann, wenn der Dienst Jobs in dem Prozess des Datenintegrationsdiensts ausführt. Wenn der Datenintegrationsdienst Jobs in separaten lokalen Prozessen oder Remoteprozessen ausführt, ignoriert der Dienst diesen Wert. Wenn Sie die Speichergröße, die der Datenintegrationsdienst zuordnen kann, nicht einschränken möchten, legen Sie diese Eigenschaft auf 0 fest. Wenn der Wert größer als 0 ist, verwendet der Datenintegrationsdienst die Eigenschaft zur Berechnung des maximalen Gesamtspeicherplatzes, der für die gleichzeitige Ausführung aller Anfragen zulässig ist. Der Datenintegrationsdienst berechnet den maximalen Gesamtspeicherplatz folgendermaßen: Maximale Speichergröße + maximale Heap-Größe + zum Laden von Programmkomponenten erforderlicher Speicherplatz Standardwert ist 0. **Hinweis:** Wenn Sie Profile oder Datenqualitäts-Mappings ausführen, müssen Sie diese Eigenschaft auf 0 festlegen.

Ergebnissatz-Cache für einen SQL-Datendienst

Wenn Sie den Ergebnissatz-Cache konfigurieren, speichert der Data Integration Service die Ergebnisse des DTM-Prozesses, der mit den einzelnen SQL-Datendienst-Abfragen und Web-Dienst-Anfragen verknüpft ist. Data Integration Service nutzt die zwischengespeicherten Daten über den gesamten Ablaufzeitraum, den Sie festlegen. Wenn ein Client dieselbe Abfrage stellt, bevor der Cache abläuft, gibt der Data Integration Service das zwischengespeicherte Ergebnis zurück.

Zum Beheben von Engpässen beim Ergebnissatz-Cache können Sie die folgende Lösung verwenden:
Konfigurieren Sie den Ergebnissatz-Cache für einen SQL-Datendienst.

Mit Hilfe des Ergebnissatz-Caches kann der Data Integration Service zwischengespeicherte Ergebnisse für SQL-Datendienst-Abfragen verwenden. Benutzer, die innerhalb eines kurzen Zeitraums identische Abfragen ausführen, können mit dem Ergebnissatz-Cache die Laufzeit identischer Abfragen verringern.

Wenn Sie dem Data Integration Service die Nutzung zwischengespeicherter Ergebnisse ermöglichen, erhöht sich die Leistung von Datendiensten. Um die Verarbeitung identischer Abfragen durch den Datendienst zusätzlich zu verkürzen, sollten Sie außerdem ausreichend Arbeitsspeicher für den Cache reservieren. Wenn Sie einen Umfang an Cache-Arbeitsspeicher festlegen, der größer oder gleich dem Umfang ist, der zum Zwischenspeichern der Ergebnisse benötigt wird, erhöhen Sie die Leistung, indem Sie den E/A-Aufwand des Systems verringern. Wenn der Data Integration Service Cache-Dateien auf den

Datenträger schreibt, erhöht sich aufgrund des E/A-Aufwands des Systems die Datendienst-Verarbeitungszeit.

Eigenschaften des Ergebnissatz-Caches für den SQL-Datendienst

Sie können die Eigenschaften des Ergebnissatz-Caches für einen Data Integration Service konfigurieren, um die Leistung zu erhöhen. Sie können außerdem die Anzahl der Millisekunden festlegen, in denen der Ergebnissatz-Cache für einen SQL-Datendienst verfügbar ist.

Die folgende Tabelle beschreibt die Eigenschaften des Ergebnissatz-Caches für den Data Integration Service:

Eigenschaft	Beschreibung
Dateinamenpräfix	Das Präfix für die Namen aller Ergebnissatz-Cachedateien, die auf dem Datenträger gespeichert sind. Standardwert ist RSCACHE.
Verschlüsselung aktivieren	Gibt an, ob die Ergebnissatz-Cachedateien mit der 128-Bit-AES-Verschlüsselung verschlüsselt werden. Gültige Werte sind „True“ oder „False“. Standardwert ist „True“.

In der folgenden Tabelle wird die Eigenschaft beschrieben, mit der die Anzahl der Millisekunden festgelegt wird, während derer der Ergebnissatz-Cache für den SQL-Datendienst verfügbar ist:

Eigenschaft	Beschreibung
Ablaufzeitraum für den Ergebnissatz-Cache	Zeitraum in Sekunden, in dem der Ergebnisdaten-Cache verwendet werden kann. Wenn -1 gesetzt ist, läuft der Ergebnisdaten-Cache nie ab. Wenn 0 gesetzt ist, ist der Ergebnisdaten-Cache deaktiviert. Änderungen des Ablaufzeitraums gelten nicht für vorhandene Caches. Wenn Sie möchten, dass alle Caches denselben Ablaufzeitraum haben, löschen Sie den Ergebnisdaten-Cache, nachdem Sie den Ablaufzeitraum geändert haben. Voreinstellung ist 0.

Aktivieren des Ergebnissatz-Caches für einen SQL-Datendienst

Um zwischengespeicherte Ergebnisse für identische SQL-Datendienst-Anfragen zu verwenden, konfigurieren Sie die Verwendung des Ergebnissatz-Caches durch den Data Integration Service.

1. Wählen Sie im Administrator Tool einen Data Integration Service aus.
2. Klicken Sie auf die Ansicht **Prozess**, um die Eigenschaften des Ergebnissatz-Caches zu konfigurieren.
3. Klicken Sie auf die Ansicht **Anwendung** und dann auf den SQL-Datendienst, um die Eigenschaft „Ablaufzeitraum für den Ergebnissatz-Cache“ zu konfigurieren.

Dauerhaft virtuelle Daten in temporären Tabellen

Eine temporäre Tabelle ist eine Tabelle in einer relationalen Datenbank, die temporäre Zwischendaten speichert. Komplexe Abfragen erfordern allgemein Speicherplatz für große Zwischendatenmengen, z. B. Informationen von Joins. Wenn Sie temporäre Tabellen implementieren, können Business Intelligence-Tools diese Daten aus der temporären Tabelle anstelle des SQL-Datendienstes abrufen. Dies führt zu einer besseren Leistung.

Temporäre Tabellen bieten außerdem auf zweierlei Weise eine höhere Sicherheit. Erstens: Nur der Benutzer der aktiven Sitzung kann auf die Tabellen zugreifen. Außerdem werden die Tabellen beibehalten, wenn eine Sitzung aktiv ist, und die Datenbank löscht die Tabellen, wenn die Verbindung geschlossen wird.

Implementierung temporärer Tabellen

Sie können temporäre Tabellen verwenden, um die Leistung von großen komplexen Abfragen zu verbessern. Temporäre Tabellen verbessern die Leistung, da Abfragen in temporären Tabellen einer relationalen Datenbank schneller als wiederholte Abfragen im SQL-Datendienst für dasselbe Dataset sind.

Die Implementierung von temporären Tabellen für Leistungsverbesserung erfordert Aktionen vom Informatica Administrator und einem Entwickler für ein Business Intelligence-Tool.

Zunächst erstellt der Informatica Administrator eine relationale Datenbankverbindung und konfiguriert den Datenintegrationsdienst zur Verwendung der Verbindung.

Der Entwickler für ein Business Intelligence-Tool (zum Beispiel IBM Cognos oder SAP Business Objects) erstellt eine Verbindung zwischen dem Business Intelligence-Tool und dem Informatica SQL-Datendienst. Die Verbindung verwendet die Informatica ODBC- oder JDBC-Treiber.

Wenn diese Verbindungen aktiv sind, kann das Business Intelligence-Tool temporäre Tabellen erstellen und verwenden, um große Mengen von temporären Daten zu verarbeiten.

KAPITEL 9

Web-Dienst-Optimierung

Dieses Kapitel umfasst die folgenden Themen:

- Web-Dienst-Optimierung - Übersicht, 80
- Optimieren von HTTP-Anfragen, 81
- Komprimierung von Web-Dienst-Meldungen, 81
- Webdienst-Optimierungsebene, 81
- Webdiensteigenschaften für Arbeitsspeicher und gleichzeitige Anfragen , 83
- Webdienst-Eigenschaft zum Konfigurieren einer aktiven DTM-Instanz, 86
- Zwischenspeichern von Web-Dienst-Ergebnissätzen, 86
- Verwaltung von Web-Dienst-Logs, 87

Web-Dienst-Optimierung - Übersicht

Sie können Web-Dienste optimieren, um die Leistung beim Ausführen von Web-Dienst-Anfragen durch den Data Integration Service zu verbessern. Passen Sie den Data Integration Service an, um Arbeitsspeicher zu verwalten und gleichzeitige Web-Dienst-Anfragen zu behandeln. Um die Web-Dienst-Leistung zu verbessern, komprimieren Sie Web-Dienst-Meldungen, optimieren Sie HTTP-Anfragen und konfigurieren Sie den Datenobjekt-Cache sowie den Ergebnissatz-Cache und konfigurieren Sie Fehler-Log-Ebenen.

Verwenden Sie zum Optimieren eines Web-Dienstes die folgenden Optimierungsverfahren:

- Optimieren Sie HTTP-Anfragen.
- Komprimieren Sie Web-Dienst-Meldungen.
- Konfigurieren Sie die Webdienst-Optimierungsebene.
- Konfigurieren Sie Web-Dienst-Eigenschaften für Parallelität und Speicher eines Data Integration Service-Prozesses.
- Legen Sie in der Konfiguration des Data Integration Service fest, dass ein DTM-Prozess aktiv bleibt, damit er mehrere Web-Dienst-Anfragen verarbeiten kann.
- Konfigurieren Sie die Datenobjekt-Zwischenspeicherung für den Webdienst.
- Konfigurieren Sie Ergebnis-Caching für die Webdienste.
- Konfigurieren Sie die Laufzeit-Fehler-Log-Ebenen der Web-Dienste.

Verwandte Themen:

- "Datenobjekt-Caching" auf Seite 66

Optimieren von HTTP-Anfragen

Optimieren Sie HTTP-Anfragen, um die Anzahl der Anfragen an den Webserver zu reduzieren.

Zum Beheben von Engpässen bei HTTP-Anfragen können Sie die folgenden Lösungen verwenden:
Verringern Sie das HTTP-Socket-Timeout für den Web-Dienst-Client.

Mit dem Socket-Timeout wird festgelegt, wie lange der Client wartet, bis ein Timeout der HTTP-Anfrage erfolgt. Bei einem großen Wert für das Socket-Timeout kann der Web-Dienst-Client abstürzen.

Komprimierung von Web-Dienst-Meldungen

Sie können die Web-Dienst-Leistung optimieren, indem Sie große Web-Meldungen, die zwischen Providern übertragen werden, komprimieren.

Berücksichtigen Sie die folgende Lösung für Engpässe bei Webdienst-Nachrichten:
Aktivieren Sie die SOAP-Meldungskomprimierung für einen Web-Dienst-Client.

Mit der SOAP-Meldungskomprimierung kann der Web-Dienst komprimierte Web-Dienst-Client-Meldungen empfangen. Der Web-Dienst kann eine SOAP-Meldung mit GZip-Komprimierung von einem Web-Dienst-Client annehmen.

Wenn der Data Integration Service die Antwort vom Web-Dienst empfängt, prüft er die Content-Encoding-HTTP-Kopfzeile in der SOAP-Meldung und dekodiert die Meldung.

Webdienst-Optimierungsebene

Der Data Integration Service optimiert Webdienste, die auf der Optimierungsebene basieren, die Sie konfigurieren. Konfigurieren Sie die Optimierungsebene, wenn Sie möchten, dass der Webdienst eine Optimierungsebene verwendet, die von der normalen Optimierungsebene abweicht. Standardmäßig verwendet jeder Webdienst die normale Optimierungsebene.

Sie können eine der folgenden Optimierungsebenen auswählen:

Keiner

Der Datenintegrationsdienst wendet keine Optimierung an.

Minimal

Der Datenintegrationsdienst wendet die frühe Projektionsoptimierungsmethode an.

Normal

Der Datenintegrationsdienst wendet folgende Optimierungsmethoden an: frühe Projektion, frühe Auswahl, Verzweigungsbereinigung, Push-Into, globale Vorhersage und Vorhersage. Normal ist die Standard-Optimierungsebene.

Komplett

Der Datenintegrationsdienst wendet folgende Optimierungsmethoden an: kostenbasiert, frühe Projektion, frühe Auswahl, Verzweigungsbereinigung, Vorhersage, Push-Into, Semi-Join und Dataship-Join.

Sie können eine oder mehrere der folgenden Methoden verwenden, um die Optimierungsebene für einen Webdienst zu konfigurieren:

- Konfigurieren Sie die Optimierungsebene für die Datenvorschau eines Webdienstes, bevor Sie ihn in einem Data Integration Service bereitstellen.
- Konfigurieren Sie die Optimierungsebene für bereitgestellte Webdienste, die auf einem bestimmten Data Integration Service ausgeführt werden.
- Konfigurieren Sie die Optimierungsebene in der Kopfzeile der Webdienstanfrage für einen bereitgestellten Webdienst.

Konfigurieren der Webdienst-Optimierungsebene für die Datenvorschau

Konfigurieren Sie die Optimierungsebene, die der Data Integration Services verwendet, um eine Vorschau der Ausgabe eines Webdienstes zu sehen.

1. Klicken Sie im Developer-Tool auf **Ausführen** > **Öffnen des ausgeführten Dialogs**.

   Das Dialogfeld **Ausführen** wird angezeigt.
2. Klicken Sie auf **Webdienst-Konfiguration**.
3. Klicken Sie auf die Schaltfläche **Neu**.
4. Geben Sie einen Namen für die Webdienst-Konfiguration ein.
5. Klicken Sie auf die Registerkarte **Erweitert**.
6. Wählen Sie eine Optimierungsebene.
7. Klicken Sie auf **Anwenden**.
8. Klicken Sie auf **Schließen**

   Das Developer-Tool erstellt die Webdienst-Konfiguration.

Wenn Sie den Daten-Viewer zur Vorschau der Ausgabe eines Vorgangs-Mappings ausführen, wählen Sie die Webdienst-Konfiguration, die eine Optimierungsebene enthält, die Sie verwenden möchten.

Konfigurieren der Optimierungsebene für bereitgestellte Web Services

Konfigurieren Sie die Optimierungsebene, die die Data Integration Services verwenden, um einen bereitgestellten Webdienst auszuführen. Sie können auswählen, die Optimierungsebene für eine einzige Anfrage durch Konfigurieren der Optimierungsebene in der HTTP-Kopfzeile der Webdienst-SOAP-Anfrage, zu überschreiben.

1. Wählen Sie im Administratortool einen Data Integration Service aus.
2. Klicken Sie auf die Ansicht **Anwendungen**.
3. Erweitern Sie die Anwendung, die den Webdienst enthält, für den Sie die Optimierungsebene konfigurieren möchten.

4. Wählen Sie den Webdienst und bearbeiten Sie die folgende Eigenschaft:

Eigenschaft	Beschreibung
Optimierungslevel	Die Optimierungsebene, die der Data Integration Service für das Objekt anwendet. Geben Sie den numerischen Wert ein, der mit der Optimierungsebene verbunden ist, die Sie konfigurieren möchten. Sie können Sie einen der folgenden numerischen Werte eingeben: - 0. Der Datenintegrationsdienst wendet keine Optimierung an. - 1. Der Datenintegrationsdienst wendet die frühe Projektionsoptimierungsmethode an. - 2. Der Datenintegrationsdienst wendet die Optimierungsmethoden „Frühe Projektion“, „Frühe Auswahl“, „Push-Into“ und „Prädikat“ an. - 3. Der Datenintegrationsdienst wendet die Optimierungsmethoden „Kostenbasiert“, „Frühe Projektion“, „Frühe Auswahl“, „Push-Into“, „Prädikat“ und „Semi-Join“ an.

5. Um die Webdienst-Optimierungsebene für eine Webdienst-Anfrage zu überschreiben, schließen Sie folgenden Eintrag in die HTTP-Kopfzeile der Webdienst-SOAP-Anfrage ein:
`WebServiceOptions.optimizeLevel= <numeric_optimizer_level>`.

Webdiensteigenschaften für Arbeitsspeicher und gleichzeitige Anfragen

Um die Webdienst-Leistung zu optimieren, konfigurieren Sie im Administrator Tool Parallelitäts- und Arbeitsspeicher-Eigenschaften für den Datenintegrationsdienst und jeden Webdienst.

Die folgende Tabelle beschreibt die Eigenschaft „Maximale Speichergröße pro Anfrage“ für das Webdienstmodul:

Eigenschaft	Beschreibung
Maximale Speichergröße pro Anfrage	Das Verhalten von „Maximale Speichergröße pro Anfrage“ richtet sich nach den folgenden Datenintegrationsdienst-Konfigurationen: - Der Dienst führt Jobs in lokalen Prozessen oder Remoteprozessen aus oder die Diensteigenschaft „Maximale Speichergröße“ lautet 0 (Standardeinstellung). „Maximale Speichergröße pro Anfrage“ ist die maximale Speichermenge in Byte, die der Datenintegrationsdienst allen Umwandlungen zuordnen kann, die den automatischen Cache-Modus in einer einzelnen Anfrage verwenden. Der Dienst weist Arbeitsspeicher separat zu Umwandlungen zu, die über eine bestimmte Cache-Größe verfügen. Der von der Anfrage verwendete Gesamtspeicher kann den Wert für „Maximale Speichergröße pro Anfrage“ überschreiten. - Der Dienst führt Jobs in dem Prozess des Datenintegrationsdiensts aus und die Diensteigenschaft „Maximale Speichergröße“ ist größer als 0. „Maximale Speichergröße pro Anfrage“ ist die maximale Speichermenge in Byte, die der Datenintegrationsdienst einer einzelnen Anfrage zuordnen kann. Der von der Anfrage verwendete Gesamtspeicher darf den Wert für „Maximale Speichergröße pro Anfrage“ nicht überschreiten. Standardwert ist 50.000.000.

In der folgenden Tabelle werden die Ausführungsoptionen für den Datenintegrationsdienst beschrieben:

Eigenschaft	Beschreibung
Maximale Ausführungspoolgröße	Maximale Anzahl an Jobs, die pro Datenintegrationsdienstprozess gleichzeitig ausgeführt werden können. Zu den Jobs gehören Datenvorschauen, Zuordnungen, Profilerstellungsjobs, SQL-Abfragen und Webdienstanfragen. Ein Datenintegrationsdienstgitter umfasst beispielsweise drei laufende Dienstprozesse. Wenn Sie den Wert auf 10 festlegen, kann jeder Datenintegrationsdienst bis zu 10 Jobs gleichzeitig ausführen. Insgesamt 30 Jobs können gleichzeitig auf dem Gitter ausgeführt werden. Standardwert ist 10. **Hinweis:** Wenn Sie den Wert für die Poolgröße erhöhen, verwendet der Datenintegrationsdienst mehr Hardware-Ressourcen wie CPU, Arbeitsspeicher und System-E/A. Legen Sie diesen Wert basierend auf den auf dem Computer verfügbaren Ressourcen fest. Berücksichtigen Sie beispielsweise die Anzahl der CPUs auf dem Computer, der den Datenintegrationsdienst hostet, und die Menge an Arbeitsspeicher, die für den Datenintegrationsdienst verfügbar ist.
Maximale Speichergröße	Die maximale Speichermenge in Byte, die der Datenintegrationsdienst für die gleichzeitige Ausführung aller Anfragen zuordnen kann, wenn der Dienst Jobs in dem Prozess des Datenintegrationsdiensts ausführt. Wenn der Datenintegrationsdienst Jobs in separaten lokalen Prozessen oder Remoteprozessen ausführt, ignoriert der Dienst diesen Wert. Wenn Sie die Speichergröße, die der Datenintegrationsdienst zuordnen kann, nicht einschränken möchten, legen Sie diese Eigenschaft auf 0 fest. Wenn der Wert größer als 0 ist, verwendet der Datenintegrationsdienst die Eigenschaft zur Berechnung des maximalen Gesamtspeicherplatzes, der für die gleichzeitige Ausführung aller Anfragen zulässig ist. Der Datenintegrationsdienst berechnet den maximalen Gesamtspeicherplatz folgendermaßen: Maximale Speichergröße + maximale Heap-Größe + zum Laden von Programmkomponenten erforderlicher Speicherplatz Standardwert ist 0. **Hinweis:** Wenn Sie Profile oder Datenqualitäts-Mappings ausführen, müssen Sie diese Eigenschaft auf 0 festlegen.

Die folgende Tabelle beschreibt die HTTP-Konfigurationseigenschaften für den Datenintegrationsdienst-Prozess:

Eigenschaft	Beschreibung
Maximale Anzahl an Backlog-Anfragen	Anzahl der HTTP- oder HTTPS-Verbindungen, die in der Warteschlange für diesen Datenintegrationsdienst-Prozess warten können. Standardwert ist 100.
Maximale Anzahl an gleichzeitigen Anfragen	Anzahl der HTTP- oder HTTPS-Verbindungen, die zu diesem Datenintegrationsdienst-Prozess hergestellt werden können. Standardwert ist 200. **Hinweis:** Für einen Webdienst wirkt sich diese Eigenschaft auf die Anzahl der Webdienstanfragen aus, die der Datenintegrationsdienst akzeptiert, bevor er sie an das Datenintegrationsdienst-Backlog sendet.

In der nachstehenden Tabelle ist die Eigenschaft „Maximale Heap-Größe“ beschrieben, die Sie für den Datenintegrationsdienst konfigurieren können:

Eigenschaft	Beschreibung
Maximale Heap-Größe	RAM-Größe für die Java Virtual Machine (JVM), auf der der Datenintegrationsdienst ausgeführt wird Mit dieser Eigenschaft verbessern Sie die Leistung. Fügen Sie einen der folgenden Buchstaben an den Wert an, um die Einheiten anzugeben: - b für Byte. - k für Kilobyte - m für Megabyte - g for gigabytes Standardwert ist 640 Megabyte. **Hinweis:** Sie können die maximale Heap-Größe erhöhen, wenn der Datenintegrationsdienst große Mengen von Daten verarbeiten muss.

Datenintegrationsdienst-Beispielkonfiguration für gleichzeitige Web-Dienst-Anfragen

Wenn Sie die Verarbeitung gleichzeitiger Web-Dienst-Anfragen durch den Datenintegrationsdienst konfigurieren, stellen Sie sicher, dass für den Web-Dienst und den Datenintegrationsdienst-Prozess die gleiche maximale Anzahl gleichzeitiger Anfragen festgelegt ist.

Beispielsweise akzeptiert der Datenintegrationsdienst in der folgenden Konfiguration 200 gleichzeitige HTTP-Anfragen, jedoch nur 10 gleichzeitige Web-Dienst-Anfragen:

Eigenschaftstyp	Eigenschaftsname	Konfiguration
Datenintegrationsdienst - Prozess	Maximale Anzahl an gleichzeitigen Anfragen	200
Datenintegrationsdienst - Prozess	Maximale Anzahl an Backlog-Anfragen	500
Datenintegrationsdienst	Maximale Ausführungspoolgröße	100
Webdienst	Maximale Anzahl an gleichzeitigen Anfragen	10

Wenn der Datenintegrationsdienst 20 Web-Dienst-Anfragen empfängt, schlagen 10 Web-Dienst-Anfragen fehl, da der Web-Dienst nur 10 gleichzeitige Anfragen akzeptiert.

Um zu verhindern, dass Web-Dienst-Anfragen fehlschlagen, wenn die maximale Anzahl gleichzeitiger Web-Dienst-Anfragen erreicht wurde, legen Sie für den Datenintegrationsdienst-Prozess und den Web-Dienst den gleichen Maximalwert fest. Wenn die Anzahl der an den Datenintegrationsdienst gesendeten Anfragen die maximale Anzahl gleichzeitiger Anfragen überschreitet, werden die überzähligen Anfragen zurückgestellt, bis sie vom Datenintegrationsdienst verarbeitet werden können.

Webdienst-Eigenschaft zum Konfigurieren einer aktiven DTM-Instanz

Zum Erhöhen der Leistung können Sie in der Konfiguration des Datenintegrationsdiensts festlegen, dass eine DTM-Instanz aktiv bleibt, damit sie mehrere Webdienst-Anfragen verarbeiten kann. Sie können im Administrator-Tool die Eigenschaft „DTM-Keep Alive-Zeit“ für den Datenintegrationsdienst konfigurieren.

Die Eigenschaft „DTM-Keep-Alive-Zeit“ ist in der folgenden Tabelle beschrieben:

Eigenschaft	Beschreibung
DTM-Keep Alive-Zeit	Anzahl der Millisekunden, für die die DTM-Instanz geöffnet bleibt, nachdem sie die letzte Anfrage abgeschlossen hat. Webdienstanfragen für denselben Vorgang können die offene Instanz wiederverwenden. Verwenden Sie die Keep Alive-Zeit, um die Leistung zu erhöhen, wenn die für die Verarbeitung der Anfrage erforderliche Zeit im Vergleich zur Dauer der Initialisierung der DTM-Instanz gering ist. Wenn die Anfrage fehlschlägt, wird die DTM-Instanz beendet. Der Standardwert ist 5000. **Hinweis:** Durch die Fähigkeit, eine bestehende DTM-Instanz zu verwenden, wird die Leistung erhöht. Der DIS benötigt zusätzliche Ressourcen, um für jede Anfrage eine DTM-Instanz zu starten. Das Aktivlassen des DTM verbraucht Arbeitsspeicher. Daher sollten Benutzer die Arbeitsspeichernutzung berücksichtigen, wenn sie diese Option konfigurieren.

Zwischenspeichern von Web-Dienst-Ergebnissätzen

Wenn Sie den Ergebnissatz-Cache konfigurieren, werden die Ergebnisse des mit den einzelnen Web-Dienst-Anfragen verbundenen DTM-Prozesses vom Data Integration Service zwischengespeichert. Der Data Integration Service nutzt die zwischengespeicherten Daten über den gesamten Ablaufzeitraum, den Sie festlegen. Wenn ein externer Client dieselbe Anfrage stellt, bevor der Cache abläuft, gibt der Data Integration Service das zwischengespeicherte Ergebnis zurück.

Zum Beheben von Engpässen beim Ergebnissatz-Cache können Sie die folgende Lösung verwenden:

Konfigurieren Sie den Ergebnissatz-Cache für einen Web-Dienst.

Mit dem Ergebnissatz-Cache kann der Data Integration Service auf zwischengespeicherte Ergebnisse für Web-Dienst-Anfragen zurückgreifen. Diese Einstellung empfiehlt sich besonders für Benutzer, die in kurzen Zeitabständen identische Abfragen senden, da damit die Laufzeit dieser Abfragen verkürzt wird.

Bei Verwendung von WS-Security erfolgt die Zwischenspeicherung der Ergebnissätze für Web-Dienste nach Benutzer. Der Data Integration Service speichert den Cache nach dem Benutzernamen, der im username-Token der Web-Dienst-Anfrage bereitgestellt wird. Beim Zwischenspeichern der Ergebnisse nach Benutzer gibt der Data Integration Service die zwischengespeicherten Ergebnisse nur an den Benutzer zurück, der die Web-Dienst-Anfrage gesendet hat.

Aktivieren des Ergebnissatz-Caches für einen Web-Dienst

Um zwischengespeicherte Ergebnisse für identische Web-Dienst-Anfragen zu verwenden, konfigurieren Sie die Verwendung des Ergebnissatz-Caches durch den Data Integration Service.

1. Wählen Sie im Administrator Tool einen Data Integration Service aus.
2. Klicken Sie auf die Ansicht **Prozess**, um die Eigenschaften des Ergebnissatz-Caches zu konfigurieren.

3. Klicken Sie auf die Ansicht **Anwendung**, klicken Sie auf den Web-Dienst, und klicken Sie dann auf die Operation, um in den Web-Dienst-Operationseigenschaften den Cache-Ablaufzeitraum zu konfigurieren. Wenn der Data Integration Service die Ergebnisse nach Benutzer zwischenspeichern soll, aktivieren Sie in den Web-Dienst-Eigenschaften WS-Security.
4. Wenn die Web-Dienst-Operation zum Zwischenspeichern des Ergebnissatzes konfiguriert ist und Sie das Zwischenspeichern des Ergebnissatzes für eine Web-Dienst-Anfrage deaktivieren möchten, schließen Sie in den HTTP-Header der SOAP-Anfrage folgende Syntax ein:

   ```
   WebServiceOptions.disableResultSetCache=true
   ```

Verwaltung von Web-Dienst-Logs

Die E/A-Leistung des Systems kann sich verringern, wenn der Data Integration Service eine große Anzahl von Log-Dateien schreibt und verwaltet. Der Data Integration Service generiert Web-Dienst-Laufzeit-Logs auf Grundlage der von Ihnen konfigurierten Tracingebene. Sie können die Anzahl der Log-Dateien steuern, die der Data Integration Service schreibt und verwaltet.

Zum Beheben von Engpässen bei Web-Dienst-Logs können Sie die folgenden Lösungen verwenden:

Deaktivieren Sie die Web-Dienst-Tracingebene.

Wenn Sie die Web-Dienst-Eigenschaften für einen bereitgestellten Web-Dienst konfigurieren, können Sie die Log-Tracingebene angeben. Die Tracingebene bestimmt, welche Typen von Logs der Data Integration Service in den Laufzeit-Log-Speicherort schreibt. Die Standard-Tracingebene für Web-Dienste lautet INFO. Wenn die Tracingebene auf SUPERFEIN oder ALLE festgelegt ist, kann sich die Leistung verringern, da der Data Integration Service mehr Logs in die Log-Datei schreibt. Die Leistungseinbuße durch Festlegen der Tracingebene auf SUPERFEIN oder ALLE ist am größten, wenn der Web-Dienst HTTPS und WS-Security verwendet.

Archivieren Sie Log-Dateien, die nicht mehr benötigt werden.

Die System-E/A wird beeinträchtigt, wenn zu viele Log-Dateien gespeichert werden. Der Data Integration Service schreibt standardmäßig Web-Dienst-Laufzeit-Logs in das folgende Verzeichnis:
`<InformaticaInstallationDir>/tomcat/bin/disLogs/ws`

Hinweis: Wenn Sie den Ordner ws löschen, um die Logs zu leeren, müssen Sie den Ordner ws neu erstellen. Beenden Sie den Data Integration Service, bevor Sie den Ordner ws löschen und neu erstellen.

KAPITEL 10

Verbindungsoptimierung

Dieses Kapitel umfasst die folgenden Themen:

- Verbindungsoptimierung - Übersicht, 88
- Verbindungspooling, 88
- Größe von Datenbank-Netzwerkpaketen, 90

Verbindungsoptimierung - Übersicht

Sie können Verbindungen optimieren, um die Leistung zu verbessern. Sie können den Pool inaktiver Verbindungsinstanzen für eine Datenbankverbindung verwalten. Sie können die Netzwerkpaketgröße erhöhen, damit größere Datenpakete das Netzwerk gleichzeitig durchlaufen können.

Verwenden Sie zum Optimieren von Verbindungen die folgenden Verfahren:

- Optimieren Sie Verbindungspooling.
- Optimieren Sie die Größe von Datenbank-Netzwerkpaketen.

Verbindungspooling

Verbindungspooling ist ein System zum Zwischenspeichern von Datenbankverbindungsinformationen, die der Data Integration Service verwendet. Es erhöht die Leistung durch Wiederverwendung zwischengespeicherter Verbindungsinformationen.

So können Sie Verbindungs-Engpässe umgehen:

Aktivieren Sie Verbindungspooling für eine Datenbankverbindung.

Aktivieren Sie Verbindungspooling, um die Verbindungsleistung zu optimieren. Sie können die inaktiven Verbindungsinstanzen für eine Datenbankverbindung verwalten. Der Verbindungspool behält inaktive Verbindungsinstanzen basierend auf den von Ihnen konfigurierten Pooling-Eigenschaften bei. Sie können die maximale und minimale Anzahl inaktiver Verbindungen sowie die maximale Wartezeit für eine inaktive Verbindung anpassen.

Poolingeigenschaften von Verbindungsobjekten

Sie können Poolingeigenschaften von Verbindungen in der Ansicht **Pooling** für eine Datenbankverbindung bearbeiten.

Die Anzahl der Verbindungspool-Bibliotheken hängt von der Anzahl der laufenden Datenintegrationsdienstprozesse oder der DTM-Prozesse ab. Jeder Datenintegrationsdienstprozess oder DTM-Prozess verwaltet eine eigene Verbindungspool-Bibliothek. Die Werte der Poolingeigenschaften gelten für jede Verbindungspool-Bibliothek.

Wenn Sie beispielsweise die maximale Anzahl von Verbindungen auf 15 einstellen, kann jede Verbindungspoolbibliothek maximal 15 inaktive Verbindungen im Pool haben. Wenn vom Datenintegrationsdienst Jobs in separaten lokalen Prozessen ausgeführt werden und drei DTM-Prozesse laufen, können maximal 45 inaktive Verbindungsinstanzen vorliegen.

Um die Gesamtanzahl inaktiver Verbindungsinstanzen zu verringern, legen Sie die Mindestanzahl an Verbindungen auf 0 fest und verringern Sie die maximal erlaubte inaktive Zeit für jede Datenbankverbindung.

Die folgende Liste beschreibt die Poolingeigenschaften der Datenbankverbindung, die Sie in der Ansicht **Pooling** für Datenbankverbindungen bearbeiten können:

Verbindungspooling aktivieren

Aktiviert das Verbindungspooling. Wenn Sie das Verbindungspooling aktivieren, behält jeder Verbindungspool inaktive Verbindungsinstanzen im Speicher. Um inaktive Verbindungen in den Pools zu löschen, müssen Sie den Datenintegrationsdienst neu starten.

Wenn das Verbindungspooling deaktiviert ist, stoppt der DTM-Prozess oder der Datenintegrationsdienst alle Poolingaktivitäten. Der DTM-Prozess oder der Datenintegrationsdienstprozess erstellt bei jeder Verarbeitung eines Jobs eine Verbindungsinstanz. Er löscht die Instanz, wenn er die Verarbeitung der Jobs beendet.

Standardwert ist aktiviert für DB2 für i5/OS-, DB2 für z/OS-, IBM DB2-, Microsoft SQL Server-, Oracle- und ODBC-Verbindungen. Die Standardeinstellung ist für Adabas-, IMS-, sequenzielle und VSAM-Verbindungen deaktiviert.

Mindestanzahl an Verbindungen

Die Mindestanzahl inaktiver Verbindungsinstanzen, die ein Pool für eine Datenbankverbindung aufrechterhält, nachdem die maximal erlaubte inaktive Zeit erreicht ist. Setzen Sie diesen Wert maximal auf die maximale Anzahl inaktiver Verbindungsinstanzen. Standardwert ist 0.

Maximale Anzahl an Verbindungen

Die maximale Anzahl inaktiver Verbindungsinstanzen, die ein Pool für eine Datenbankverbindung aufrechterhält, bevor die maximale inaktive Zeit erreicht ist. Legen Sie diesen Wert auf eine höhere Anzahl als die Mindestanzahl an inaktiven Verbindungsinstanzen fest. Standardwert ist 15.

Maximale Leerlaufzeit

Die Anzahl der Sekunden, die eine Verbindungsinstanz, welche die Mindestanzahl von Verbindungsinstanzen überschritten hat, inaktiv bleiben kann, bevor sie vom Verbindungspool gelöscht wird. Der Verbindungspool ignoriert die inaktive Zeit, wenn die Verbindungsinstanz die Mindestanzahl von inaktiven Verbindungsinstanzen nicht überschreitet. Standardwert ist 120.

Größe von Datenbank-Netzwerkpaketen

Wenn Sie in bzw. aus Oracle-, Sybase ASE- oder Microsoft SQL Server-Targets schreiben oder lesen müssen, können Sie die Leistung verbessern, indem Sie die Netzwerkpaketgröße entsprechend der Datenbank erhöhen, aus der Sie lesen oder in die Sie schreiben. Wenn Sie die Netzwerkpaketgröße erhöhen, können größere Datenpakete das Netzwerk gleichzeitig durchlaufen.

Zum Beheben von Engpässen bei Datenbank-Netzwerkpaketen können Sie die folgenden Lösungen verwenden:

Erhöhen Sie die Datenbank-Netzwerkpaketgröße für eine Oracle-Datenbank.

Sie können die Datenbank-Netzwerkpaketgröße in listener.ora und tnsnames.ora erhöhen. Weitere Informationen zum Erhöhen der Paketgröße finden Sie ggf. in der Datenbankdokumentation.

Erhöhen Sie die Datenbank-Netzwerkpaketgröße für eine Sybase ASE-Datenbank.

Informationen zum Erhöhen der Paketgröße finden Sie ggf. in der Datenbankdokumentation. Sie müssen außerdem die Paketgröße für Sybase ASE im relationalen Verbindungsobjekt im Data Integration Service entsprechend der Datenbankserver-Paketgröße ändern.

Erhöhen Sie die Datenbank-Netzwerkpaketgröße für eine Microsoft SQL Server-Datenbank.

Informationen zum Erhöhen der Paketgröße finden Sie ggf. in der Datenbankdokumentation. Sie müssen außerdem die Paketgröße für Microsoft SQL Server im relationalen Verbindungsobjekt im Data Integration Service entsprechend der Datenbankserver-Paketgröße ändern.

INDEX

"Early selection"-Optimierung
SQL-Umwandlung 35
Web-Dienst-Verbraucher-Umwandlung 39

A

Abfrageansicht
Konfigurieren von Hinweisen 22
Abfrageoptimierung
Quelloptimierung 19
Aggregat-Umwandlung
Umwandlungsoptimierung 25
Aktive DTM-Instanz
Web-Dienst 86
Aktivieren des Ergebnissatz-Caches für einen SQL-Datendienst
Ergebnissatz-Cache 78
Analyst Service-Optimierung
Laufzeitoptimierung 62

B

Bedingungsfilter
Quelloptimierung 20

D

Data Integration Service
Web-Dienst-Ergebnissatz-Cache 86
Dataship-Join-Optimierung
Beschreibung 46
Datenbanken
Optimieren von Quellen für die Partitionierung 58
Optimieren von Zielen für die Partitionierung 58
Datenbankhinweise
Eingabe im Developer Tool 22
Datenbankprüfpunkt-Intervalle
Target-Optimierung 16
Datenintegrationsdienst
Ergebnissatz-Cache für SQL-Datendienst 77
Datenintegrationsdienst-Optimierung
Laufzeitoptimierung 62
Datenobjekt-Cache
Benutzerverwaltete Tabellen 67
Beschreibung 67
Index-Cache 67
konfigurieren 67
Optimierung 69
Tabellen-Datentypen 67

E

Einfachdateien
Optimieren von Quellen für die Partitionierung 57
Optimieren von Zielen für die Partitionierung 57
Einfachdateiquelle
Quelloptimierung 19
Einfachdateiziel
Ziel-Optimierung 15
Einschränkungen
Konfigurieren von Einschränkungen 23
Quelloptimierung 22
Eliminierung von Umwandlungsfehlern
Umwandlungsoptimierung 37
Engpässe
unter UNIX 13
unter Windows 12
Ergebnissatz-Cache
Aktivieren des Ergebnissatz-Caches für einen SQL-Datendienst 78
Ergebnissatz-Cache - Eigenschaften 78
Ergebnissatz-Cache - Eigenschaften
Laufzeitoptimierung 78
Ergebnissatz-Cache für SQL-Datendienst
Datenintegrationsdienst 77
Expressionsoptimierung
Mapping-Optimierung 26

F

Fehler-Tracing
Mapping-Optimierung 54
Fehler-Tracingebene
Verwaltung von Web-Dienst-Logs 87
Filteroptimierung
Mapping-Optimierung 53
Fremdanbieter-Client-Tools
Laufzeitoptimierung 72
frühe Auswahloptimierung
Beschreibung 48
frühe Projektionsoptimierung
Beschreibung 44

G

Gleichzeitige Anfragen
SQL-Datendienst 76
Webdienst 83
Größe von Datenbank-Netzwerkpaketen
Verbindungsoptimierung 90

H

Hat Nebenwirkungen
  Umwandlungseigenschaft - Beschreibung 38
Hinweise
  Abfrageansicht 22

J

Java-Umwandlung
  Umwandlungsoptimierung 28
JDBC-Treiber
  Laufzeitoptimierung 72
Joiner-Umwandlung
  Umwandlungsoptimierung 31

K

Komprimierung von Web-Dienst-Meldungen
  Web-Dienst-Optimierung 81
konfiguriertes Datenobjekt
  Quelloptimierung 23
kostenbasierte Optimierung
  Beschreibung 46

L

Laufzeitoptimierung
  Analyst Service-Optimierung 62
  Datenintegrationsdienst-Optimierung 62
  Optimierung des Modell-Repository Service 63
  Systemoptimierung 69
  Überwachungsstatistiken 63
Leistungsoptimierung
  Dataship-Join-Optimierungsmethode 46
  frühe Auswahloptimierungsmethode 48
  frühe Projektionsoptimierungsmethode 44
  Globale Vorhersageoptimierungsmethode 49
  kostenbasierte Optimierungsmethode 46
  Methode zur Optimierung der Verzweigungsbereinigung 49
  Optimierungsebenen 43
  Optimierungsmethoden 43
  Push-Into-Optimierungsmethode 49
  Pushdown-Optimierungsmethode 50
  Semi-Join Optimierungsmethode 47
  Vorhersageoptimierungsmethode 45
Lesen in einem einzigen Durchlauf
  Mapping-Optimierung 52
Logische Datenobjekte
  Caching in der Datenbank 67
Lookup-Umwandlung
  Umwandlungsoptimierung 32

M

Mapping-Optimierung
  Expressionsoptimierung 26
  Fehler-Tracing 54
  Filteroptimierung 53
  Lesen in einem einzigen Durchlauf 52
  Optimierung der Datentypkonvertierung 53
Mappings
  Globale Vorhersageoptimierungsmethode 49
  Optimierungsmethoden 43
Mappings *(Fortsetzung)*
  partitionierte Optimierung 55
  Vorhersageoptimierungsmethode 45
Massenladevorgänge
  Target-Optimierung 16
Maximaler Parallelismus
  Vergrößern 56
Minimale Optimierungsebene
  Beschreibung 43

N

Nebeneffekte
  SQL-Umwandlung 35
  Web-Dienst-Verbraucher-Umwandlung 39
Nebenwirkungen
  Beschreibung 38
Normale Optimierungsebene
  Beschreibung 43

O

Optimieren von HTTP-Anfragen
  Web-Dienst-Optimierung 81
Optimierung
  Dataship-Join-Optimierungsmethode 46
  frühe Auswahloptimierungsmethode 48
  frühe Projektionsoptimierungsmethode 44
  kostenbasierte Optimierungsmethode 46
  Methode zur Optimierung der Verzweigungsbereinigung 49
  Methoden zur Optimierung der Mapping-Leistung 43
  Nebenwirkungen 38
  Push-Into-Optimierungsmethode 49
  Pushdown-Optimierungsmethode 50
  Semi-Join Optimierungsmethode 47
Optimierung der Datentypkonvertierung
  Mapping-Optimierung 53
Optimierung der Verzweigungsbereinigung
  Beschreibung 49
Optimierung des Modell-Repository Service
  Laufzeitoptimierung 63
Optimierungsebenen
  Beschreibung 43
Oracle-Datenbankoptimierung
  Quelloptimierung 24
  Ziel-Optimierung 17

P

Partitionierung
  mehrere CPUs 56
  optimieren 55
  Optimieren der Quelldatenbanken 58
  Optimieren von Einfachdateiquellen 57
  Optimieren von Einfachdateizielen 57
  Optimieren von Umwandlungen 59
  Optimieren von Zieldatenbanken 58
Port filtern
  Web-Dienst-Verbraucher-Umwandlung 40
Push-Into-Optimierung
  Beschreibung 49
  In der SQL-Umwandlung aktivieren 36
  SQL-Umwandlung 36
  Web-Dienst-Verbraucher-Umwandlung 40

Pushdown-Optimierung
Beschreibung 50
Pushdown-Optimierungsmethode
Quell-Pushdown 51
vollständige Übertragung 50

Q

Quelloptimierung
Abfrageoptimierung 19
Bedingungsfilter 20
Einfachdateiquelle 19
Einschränkungen 22
konfiguriertes Datenobjekt 23
Oracle-Datenbankoptimierung 24
Select Distinct 20

S

Select Distinct
Quelloptimierung 20
Semi-Join-Optimierung
Beschreibung 47
Sortierer-Umwandlung
Umwandlungsoptimierung 34
Speicherzuordnung
Aktive DTM-Instanz 86
Gleichzeitige Anfragen 76
SQL-Datendienst 76
Web-Dienst 86
SQL-Abfragepläne
anzeigen 75
SQL-Datendienst
Speicherzuordnung 76
SQL-Datendienstoptimierung
Fremdanbieter-Client-Tools 72
JDBC-Treiber 72
SQL-Hinweise
Eingabe im Developer Tool 22
SQL-Umwandlung
"Early Selection"-Optimierung 35
Eigenschaften der Push-Into-Optimierung 36
Push-Into-Optimierung 36
Umwandlungsoptimierung 35
System
Engpässe unter UNIX, bestimmen 13
Engpässe unter Windows, bestimmen 12
Systemoptimierung
Laufzeitoptimierung 69

T

Target-Optimierung
Datenbankprüfpunkt-Intervalle 16
Massenladevorgänge 16
Temporäre Tabellen
Beschreibung 78

U

Überwachungsstatistiken
Laufzeitoptimierung 63
Umwandlungen
Optimieren für die Partitionierung 59
Umwandlungs-Cache
Umwandlungsoptimierung 37
Umwandlungsoptimierung
Aggregat-Umwandlung 25
Eliminierung von Umwandlungsfehlern 37
Java-Umwandlung 28
Joiner-Umwandlung 31
Lookup-Umwandlung 32
Sortierer-Umwandlung 34
SQL-Umwandlung 35
Umwandlungs-Cache 37
Webdienstverbraucher-Umwandlung 39
UNIX
Systemengpässe 13

V

Verbindungsoptimierung
Größe von Datenbank-Netzwerkpaketen 90
Verbindungspooling 88
Verbindungspooling
Eigenschaften 89
Verbindungsoptimierung 88
Verwaltung von Web-Dienst-Logs
Fehler-Tracingebene 87
Virtuelle Tabellen
Caching in der Datenbank 67
Vollständige Optimierungsebene
Beschreibung 43

W

Web-Dienst
Speicherzuordnung 86
Web-Dienst-Ergebnissatz-Cache
Data Integration Service 86
Web-Dienst-Optimierung
Komprimierung von Web-Dienst-Meldungen 81
Optimieren von HTTP-Anfragen 81
Web-Dienst-Verbraucher-Umwandlung
"Early selection"-Optimierung 39
Filteroptimierung 40
Push-Into-Optimierung 40
Push-Into-Optimierung aktivieren 41
Webdienst
Gleichzeitige Anfragen 83
Webdienstverbraucher-Umwandlung
Umwandlungsoptimierung 39
Windows
Engpässe 12

Z

Ziel-Optimierung
Einfachdateiziel 15
Oracle-Datenbankoptimierung 17