



Informatica®

10.4.0

Guía de ajuste del rendimiento

© Copyright Informatica LLC 2009, 2020

Este software y la documentación se proporcionan exclusivamente en virtud de un acuerdo de licencia independiente que contiene restricciones de uso y divulgación. Ninguna parte de este documento puede ser reproducida o transmitida de cualquier forma o manera (electrónica, fotocopia, grabación o mediante otros métodos) sin el consentimiento previo de Informatica LLC.

Informatica y el logotipo de Informatica son marcas comerciales o marcas comerciales registradas de Informatica LLC en Estados Unidos y en las diversas jurisdicciones de todo el mundo. La lista actual de marcas comerciales de Informatica está disponible en Internet en <https://www.informatica.com/trademarks.html>. Otros nombres de productos y empresas pueden ser nombres o marcas comerciales de sus respectivos titulares.

Las bases de datos, el software y los programas de DERECHOS DEL GOBIERNO DE LOS ESTADOS UNIDOS, y la documentación e información técnica relacionadas entregadas a los clientes del Gobierno de los Estados Unidos constituyen "software informático comercial" o "datos técnicos comerciales" de acuerdo con el Reglamento de Adquisición Federal y las regulaciones complementarias específicas del organismo que correspondan. Como tales, el uso, la duplicación, la divulgación, la modificación y la adaptación están sujetos a las restricciones y los términos de licencia establecidos en el contrato gubernamental aplicable, y hasta donde sea aplicable en función de los términos del contrato gubernamental, a los derechos adicionales establecidos en FAR 52.227-19, Licencia de Software Informático Comercial.

Las partes de este software o la documentación están sujetas a derechos de autor de terceros. Se incluyen con el producto los avisos obligatorios de terceros.

La información contenida en esta documentación está sujeta a cambios sin previo aviso. Si encuentra algún problema en esta documentación, escríbanos a infa_documentation@informatica.com para notificarnoslo.

Los productos de Informatica gozan de garantía en función de los términos y condiciones de los acuerdos conforme a los cuales se proporcionen. INFORMATICA PROPORCIONA LA INFORMACIÓN DE ESTE DOCUMENTO "TAL CUAL" SIN GARANTÍA DE NINGÚN TIPO, EXPRESA O IMPLÍCITA, INCLUIDAS LAS GARANTÍAS DE COMERCIALIZACIÓN, ADAPTACIÓN A UN FIN PARTICULAR Y CUALQUIER GARANTÍA O CONDICIÓN DE NO INCUMPLIMIENTO.

Fecha de publicación: 2020-02-03

Tabla de contenido

Prefacio	7
Recursos de Informatica	7
Informatica Network.	7
Base de conocimiento de Informatica.	7
Documentación de Informatica	7
Matrices de disponibilidad de producto de Informatica.	8
Informatica Velocity.	8
Catálogo de soluciones de Informatica.	8
Servicio internacional de atención al cliente de Informatica.	8
 Capítulo 1: Resumen del ajuste del rendimiento.....	9
Resumen del ajuste del rendimiento.	9
Proceso de ajuste del rendimiento.	10
Cuellos de botella de destino.	10
Cuellos de botella de origen.	11
Cuellos de botella de asignación.	11
Cuellos de botella del sistema del equipo.	11
Identificación de cuellos de botella del sistema en Windows.	11
Identificar cuellos de botella del sistema en UNIX.	12
Cuellos de botella en tiempo de ejecución.	12
Cuellos de botella de optimización del servicio de datos SQL.	13
Cuellos de botella de optimización del servicio web.	13
Cuellos de botella de conexión.	13
 Capítulo 2: Optimización de destino.....	14
Resumen de la optimización de destino.	14
Optimización de destino de archivo sin formato.	14
Intervalos de puntos de comprobación de base de datos.	15
Cargas masivas.	15
Optimización de destino de la base de datos.	16
 Capítulo 3: Optimización de origen.....	17
Resumen de optimización del origen.	17
Optimización de origen de archivo sin formato.	18
Optimización de la consulta.	18
Filtros condicionales.	19
Seleccionar distinto.	19
Sugerencias.	20
Reglas y directrices de sugerencias.	20
Cómo crear sugerencias.	21

Restricciones.	21
Configurar restricciones.	22
Optimización del objeto de datos personalizado.	22
Optimización de origen de base de datos.	23
Capítulo 4: Optimización de transformación.	24
Optimización de transformación.	24
Optimización de la transformación de agregación.	24
Optimización de la expresión.	25
Optimización de transformación Java.	27
Optimización de primera selección con la transformación de Java.	27
Optimización de inserción con la transformación de Java.	29
Optimización de la transformación de combinación.	30
Optimización de la transformación de búsqueda.	30
Optimización de la transformación de ordenación.	33
Optimización de la transformación de SQL.	34
Optimización de primera selección con la transformación de SQL.	34
Optimización de inserción con la transformación de SQL.	34
Memoria caché de la transformación.	35
Eliminación del error de transformación.	36
Efectos secundarios de la transformación.	37
Optimización de la transformación del consumidor de servicio web.	37
Optimización de primera selección con la transformación del consumidor de servicio web.	38
Optimización de inserción con la transformación del consumidor de servicio web.	39
Capítulo 5: Optimización de la asignación.	41
Resumen de optimización de la asignación.	41
Métodos de optimización.	42
Niveles del optimizador.	42
Optimizaciones de filtro.	43
Método de optimización de primera proyección.	43
Método de optimización de predicado.	44
Método de optimización basado en el coste.	45
Método de optimización dataship-join.	46
Método de optimización semi-join.	47
Método de optimización de primera selección.	48
Método de optimización de predicado global.	48
Método de optimización de eliminación de ramificaciones.	49
Método de la optimización de inserción.	49
Optimización de inserciones.	49
Optimización de inserciones completa.	50
Inserción de origen.	51
Reglas y directrices de la optimización de inserciones.	51

Lectura de pase individual.	52
Optimización del filtro.	52
Optimización de conversión de tipos de datos.	53
Error de seguimiento.	53
Capítulo 6: Optimización de asignación con particiones.	55
Resumen de optimización de asignación con particiones.	55
Utilizar varias CPU.	56
Aumentar el número máximo de paralelismos.	56
Optimizar archivos sin formato para particiones.	57
Optimizar orígenes de archivos sin formato para particiones.	57
Optimizar destinos de archivos sin formato para particiones.	57
Optimizar las bases de datos relacionales para la partición.	58
Optimizar la base de datos de origen para la partición.	58
Optimizar la base de datos de destino para la partición.	58
Optimizar transformaciones para particiones.	59
Capítulo 7: Optimización del tiempo de ejecución.	61
Resumen de optimización del tiempo de ejecución.	61
Optimización del servicio de aplicación.	61
Optimización del servicio del analista.	62
Optimización del servicio de integración de datos.	62
Optimización del servicio de repositorio de modelos.	63
Estadísticas de supervisión.	64
Asignación de memoria.	65
Memoria caché de objetos de datos.	66
Tipos de datos de tablas de la memoria caché.	67
Optimización de la memoria caché de objetos de datos.	68
Optimización del sistema.	69
Capítulo 8: Optimización del servicio de datos SQL.	70
Resumen de la optimización del servicio de datos SQL.	70
Optimización de la herramienta de cliente de terceros.	71
Niveles del optimizador del servicio de datos SQL.	71
Configuración del nivel del optimizador del servicio de datos SQL para la vista previa de datos	72
Configuración del nivel del optimizador para los servicios de datos SQL implementados.	73
Plan de consulta del servicio de datos SQL.	73
Cómo ver un plan de consulta SQL.	74
Propiedades del servicio de datos SQL para la memoria y las solicitudes simultáneas.	75
Memoria caché del conjunto de resultados para un servicio de datos SQL.	76
Propiedades de la memoria caché del conjunto de resultados del servicio de datos SQL.	77
Cómo habilitar la memoria caché del conjunto de resultados para un servicio de datos SQL.	77

Conservación de datos virtuales en tablas temporales.	78
Implementación de tablas temporales.	78
Capítulo 9: Optimización del servicio web.	79
Resumen de optimización del servicio web.	79
Optimizar solicitudes HTTP.	80
Compresión de mensajes del servicio web.	80
Nivel del optimizador del servicio web.	80
Configuración del nivel del optimizador del servicio web para la vista previa de datos	81
Configuración del nivel del optimizador para servicios web implementados.	81
Propiedades de los servicios web para la memoria y para solicitudes simultáneas	82
Ejemplo de configuración del Servicio de integración de datos para solicitudes de servicio web simultáneas	84
Propiedad del servicio web para configurar una instancia del DTM activa.	84
Almacenamiento en memoria caché del conjunto de resultados de un servicio web.	85
Cómo habilitar la memoria caché del conjunto de resultados para un servicio web.	85
Administración del registro del servicio web.	85
Capítulo 10: Optimización de las conexiones.	87
Resumen de la optimización de las conexiones.	87
Agrupación de conexiones.	87
Propiedades de agrupación en objetos de conexión.	88
Tamaño del paquete de red de la base de datos.	89
Índice.	90

Prefacio

Consulte la *Guía de ajuste del rendimiento de Informatica®* para obtener información sobre cómo optimizar el rendimiento de las asignaciones. Sepa cómo detectar y eliminar cuellos de botella de rendimiento dentro de cada componente de asignación.

Recursos de Informatica

Informatica proporciona una variedad de recursos de productos a través de Informatica Network y otros portales en línea. Use los recursos para sacar el mayor provecho de los productos y las soluciones de Informatica y aprender de otros expertos en la materia y usuarios de Informatica.

Informatica Network

Informatica Network es la puerta de entrada a muchos recursos, entre ellos, la base de conocimientos de Informatica y el servicio internacional de atención al cliente de Informatica. Para entrar en Informatica Network, visite <https://network.informatica.com>.

Como miembro de Informatica Network, tiene las siguientes opciones:

- Buscar recursos de productos en la base de conocimientos
- Ver la información de disponibilidad del producto
- Crear y revisar casos de soporte
- Buscar su red de grupos de usuarios de Informatica locales y colaborar con sus pares

Base de conocimiento de Informatica

Use la base de conocimientos de Informatica para encontrar recursos de productos como artículos prácticos, procedimientos recomendados, tutoriales de video y respuestas a preguntas frecuentes.

Para buscar en la base de conocimiento, visite <https://search.informatica.com>. Si tiene preguntas, comentarios o ideas relacionadas con la base de conocimiento de Informatica, póngase en contacto con el equipo de la base de conocimiento de Informatica en KB_Feedback@informatica.com.

Documentación de Informatica

Use el portal de documentación de Informatica para recorrer una extensa biblioteca de documentación para las versiones de productos actuales y recientes. Para recorrer el portal de documentación, visite <https://docs.informatica.com>.

Si tiene preguntas, comentarios o ideas acerca de la documentación de los productos, póngase en contacto con el equipo de la documentación de Informatica en infa_documentation@informatica.com.

Matrices de disponibilidad de producto de Informatica

Las matrices de disponibilidad de producto (PAM, Product Availability Matrixes) indican las versiones de sistemas operativos, bases de datos y otros tipos de orígenes y destinos de datos admitidos por la versión de un producto. Puede recorrer las PAM de Informatica en <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

Informatica Velocity es una colección de consejos y procedimientos recomendados desarrollados por los servicios profesionales de Informatica que se basan en experiencias reales de cientos de proyectos de administración de datos. Informatica Velocity representa el conocimiento colectivo de los consultores de Informatica que trabajan con organizaciones de todo el mundo para planificar, desarrollar, implementar y dar mantenimiento a soluciones de administración de datos exitosas.

Puede encontrar recursos de Informatica Velocity en <http://velocity.informatica.com>. Si tiene alguna pregunta, comentario o idea acerca de Informatica Velocity, póngase en contacto con los servicios profesionales de Informatica en ips@informatica.com.

Catálogo de soluciones de Informatica

El catálogo de soluciones de Informatica es un foro donde puede buscar soluciones que aumenten, amplíen o mejoren sus implementaciones de Informatica. Aproveche cualquiera de los cientos de soluciones de socios y desarrolladores de Informatica que se encuentran en el catálogo para mejorar su productividad y acelerar la implementación de los proyectos. Puede encontrar el catálogo de soluciones de Informatica en <https://marketplace.informatica.com>.

Servicio internacional de atención al cliente de Informatica

Puede ponerse en contacto con un centro de atención global por teléfono o a través del Informatica Network.

Para encontrar el número de teléfono local del servicio internacional de atención al cliente de Informatica, visite el sitio web de Informatica en el siguiente vínculo:

<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>.

Para encontrar recursos de soporte en línea en Informatica Network, visite <https://network.informatica.com> y seleccione la opción eSupport.

CAPÍTULO 1

Resumen del ajuste del rendimiento

Este capítulo incluye los siguientes temas:

- [Resumen del ajuste del rendimiento, 9](#)
- [Proceso de ajuste del rendimiento, 10](#)
- [Cuellos de botella de destino, 10](#)
- [Cuellos de botella de origen, 11](#)
- [Cuellos de botella de asignación, 11](#)
- [Cuellos de botella del sistema del equipo, 11](#)
- [Cuellos de botella en tiempo de ejecución, 12](#)
- [Cuellos de botella de optimización del servicio de datos SQL, 13](#)
- [Cuellos de botella de optimización del servicio web, 13](#)
- [Cuellos de botella de conexión, 13](#)

Resumen del ajuste del rendimiento

El objetivo de ajustar el rendimiento es eliminar los cuellos de botella del rendimiento. Un cuello de botella es un área de la asignación que se ejecuta con más frecuencia y tiene el rendimiento más bajo. Un cuello de botella reduce el rendimiento general de la asignación.

Para optimizar una asignación, identifique un cuello de botella de rendimiento, elimínelo y, a continuación, identifique el siguiente cuello de botella de rendimiento. Optimice un componente de la asignación cada vez. Puede calcular una asignación antes y después cambiarla para comprobar que la optimización tiene un impacto en el rendimiento.

Puede encontrar y corregir cuellos de botella en los componentes de asignación, como orígenes, destinos y conexiones. Puede comprobar los cuellos de botella en el servicio de integración de datos y la máquina en la que se ejecuta el servicio de integración de datos. También puede ajustar las propiedades para los servicios web y los servicios de datos SQL.

Proceso de ajuste del rendimiento

Puede seguir un conjunto de pasos para ajustar los componentes de asignación y aumentar el rendimiento.

Puede optimizar los componentes de la asignación en el siguiente orden:

1. Destinos
2. Orígenes
3. Asignaciones
4. Transformaciones
5. Entorno de Informatica en Administrator Tool
6. El sistema del equipo
7. Servicio de datos o servicio web

Utilice los siguientes métodos para identificar los cuellos de botella de rendimiento:

- Ejecute asignaciones de prueba. Puede configurar una asignación de prueba para leer desde un origen de archivo sin formato o para escribir en un destino de archivo sin formato para identificar los cuellos de botella de origen y de destino.
- Analice los detalles del rendimiento. Analice los detalles del rendimiento, tales como métodos de optimización, para determinar donde disminuye el rendimiento de la asignación.
- Supervise el rendimiento del sistema. Puede utilizar herramientas de supervisión del sistema para ver el porcentaje de uso de CPU, las esperas de E/S, la paginación y el uso de los recursos del sistema.

Cuellos de botella de destino

Los cuellos de botella de destino son reducciones en el rendimiento cuando el servicio de integración de datos escribe en un destino. Los cuellos de botella de destino pueden ocurrir cuando la base de datos utiliza pequeños intervalos de puntos de comprobación o tamaños pequeños de paquetes de red de base de datos.

El cuello de botella de rendimiento más común se produce cuando el servicio de integración de datos escribe en una base de datos de destino. Si la base de datos utiliza pequeños intervalos de puntos de comprobación, el procesamiento de la base de datos se reduce con más frecuencia al escribir un punto de comprobación. Los tamaños pequeños de paquetes de red de base de datos pueden provocar cuellos de botella. Puede permitir paquetes de datos más grandes para cruzar la red al mismo tiempo.

Para identificar un cuello de botella de destino, puede crear una copia de la asignación que tiene un destino de archivo sin formato en lugar de un destino de base de datos. Si el rendimiento aumenta significativamente, tiene un cuello de botella de destino. Si la asignación ya escribe en un destino de archivo sin formato, probablemente no tiene un cuello de botella de destino.

Cuellos de botella de origen

Los cuellos de botella de origen son reducción del rendimiento cuando el servicio de integración de datos lee desde una base de datos de origen. Los cuellos de botella de origen pueden ocurrir cuando la consulta de origen no es eficaz o cuando los tamaños de los paquetes de red de la base de datos son pequeños.

Cuando la asignación lee desde un origen relacional, puede utilizar los siguientes métodos para identificar los cuellos de botella de origen:

- Añada una transformación de filtro a la asignación. Añada la transformación de filtro después del origen. Configure la condición de filtro en FALSE para que la transformación de filtro no devuelva ningún dato. Si la cantidad de tiempo que lleva la asignación es más o menos el mismo, la asignación tiene un cuello de botella de origen.
- Cree una asignación de prueba de lectura. Haga una copia de la asignación pero quite todas las transformaciones, combinaciones o consultas. Conecte el origen a un destino. Si el rendimiento de la asignación es similar a la asignación original, tiene un cuello de botella de origen.
- Ejecute la consulta de lectura directamente en la base de datos de origen. Copie la consulta de lectura desde el registro de asignación. Ejecute la consulta en la base de datos de origen con una herramienta de consultas como isql. Calcule el tiempo de ejecución y el tiempo que tarda la consulta en devolver una fila.

Cuellos de botella de asignación

Si se determina que no tiene un cuello de botella de origen o destino, puede que tenga un cuello de botella de asignación. Si el tamaño de la memoria caché es reducido, la memoria de búfer es baja y los intervalos de confirmación son pequeños puede provocar cuellos de botella de asignación.

Para identificar un cuello de botella de asignación, analice los detalles del rendimiento en el registro de asignación. Los detalles de rendimiento incluyen información sobre cada transformación, tales como el número de filas de entrada, filas de salida y filas de error.

También puede añadir una transformación de filtro antes de cada definición de destino. Configure la condición de filtro en FALSE para que la transformación de filtro no cargue ningún dato en las tablas de destino. Si el tiempo que lleva ejecutar la nueva asignación es el mismo que para la asignación original, tiene un cuello de botella de asignación.

Cuellos de botella del sistema del equipo

Puede ver el uso de los recursos cuando ejecute servicios de Informática en Windows o UNIX. En Windows utilice el Administrador de tareas. UNIX tiene varias herramientas que puede utilizar para revisar el rendimiento.

Identificación de cuellos de botella del sistema en Windows

Puede ver las fichas de rendimiento y procesos en el Administrador de tareas para obtener información del sistema. La ficha rendimiento del administrador de tareas proporciona un resumen del uso de CPU y la cantidad total de memoria utilizada. Utilice el supervisor de rendimiento para ver más información detallada.

La siguiente tabla describe la información del sistema que puede utilizar en el supervisor de rendimiento de Windows para crear un gráfico:

Propiedad	Descripción
Tiempo del procesador de porcentaje	Si tiene más de una CPU, supervise cada una para obtener el tiempo del procesador de porcentaje.
Páginas/segundo	Si las páginas/segundo es mayor que cinco, puede que tenga demasiada presión de memoria, conocida como hiperpaginación.
Tiempo de porcentaje de los discos físicos	El porcentaje de tiempo que el disco físico está ocupado realizando lecturas o escribiendo solicitudes.
Longitud de cola de discos físicos	El número de usuarios en espera para acceder al mismo dispositivo de disco.
Bytes totales del servidor por segundo	El servidor ha enviado y recibido desde la red.

Identificar cuellos de botella del sistema en UNIX

Utilice las siguientes herramientas para identificar cuellos de botella del sistema en UNIX:

- **top.** Ver el rendimiento general del sistema. Esta herramienta muestra el uso de CPU, el uso de la memoria y el uso de intercambio para el sistema y para procesos individuales que se ejecutan en el sistema.
- **iostat.** Supervisar la operación de carga de cada disco adjunto al servidor de la base de datos. **iostat** muestra el porcentaje de tiempo que el disco está físicamente activo. Si utiliza matrices de disco, haga uso de las utilidades proporcionadas con las matrices de disco en lugar de **iostat**.
- **vmstat.** Supervisar las acciones de intercambio de disco.
- **sar.** Ver informes de la actividad del sistema detallados de uso de CPU, memoria y disco. Puede usar esta herramienta para supervisar la carga de CPU. Proporciona el uso en porcentaje del usuario, sistema, tiempo de inactividad y tiempo de espera. También puede usar esta herramienta para supervisar las acciones de intercambio de disco.

Cuellos de botella en tiempo de ejecución

Habilite las características de rendimiento y ajuste las propiedades de Servicio de integración de datos para optimizar el rendimiento de las asignaciones. Configure la optimización para el servicio de integración de datos y el servicio de repositorio de modelos en la herramienta Administrator.

Asigne memoria para obtener un rendimiento del sistema óptimo y configure los niveles de seguimiento de errores para reducir el número de eventos de registro generados por el servicio de integración de datos cuando ejecuta la asignación.

Puede configurar la cantidad máxima de memoria que el servicio de integración de datos asigna para ejecutar todas las solicitudes simultáneas. Puede configurar la cantidad máxima de memoria que el servicio de integración de datos asigna para cualquier solicitud dada.

Puede configurar la memoria caché del conjunto de resultados para permitir que el servicio de integración de datos almacene en la memoria caché los resultados de los procesos DTM asociados con cada consulta del servicio de datos SQL y solicitud de servicio web.

Cuellos de botella de optimización del servicio de datos SQL

Puede optimizar los servicios de datos SQL para mejorar el rendimiento cuando los usuarios finales ejecutan consultas SQL en los mismos mediante herramientas de cliente de terceros. Si un servicio de datos SQL utiliza una asignación de tabla virtual, puede optimizar las transformaciones y la asignación.

Puede optimizar el controlador JDBC para mejorar el rendimiento cuando consulte un servicio de datos SQL. También puede configurar la memoria caché del objeto de datos para el servicio de integración de datos y mejorar el rendimiento de las asignaciones y consultas SQL.

Cuellos de botella de optimización del servicio web

Puede optimizar los servicios web para mejorar el rendimiento cuando el servicio de integración de datos ejecuta solicitudes de servicio web. Ajuste el servicio de integración de datos para administrar la memoria, administrar solicitudes simultáneas de servicio web y mantener un proceso DTM activo de modo que puede procesar más de una solicitud de servicio web.

Para mejorar el rendimiento del servicio web, utilice la compresión de mensajes del servicio web, optimice las solicitudes HTTP y configure la memoria caché del objeto de datos.

Cuellos de botella de conexión

Puede optimizar las conexiones para mejorar el rendimiento. Puede administrar el grupo de instancias de conexiones inactivas para una conexión de base de datos. Puede aumentar el tamaño de paquetes de red para permitir que paquetes de mayor tamaño puedan cruzar la red al mismo tiempo.

CAPÍTULO 2

Optimización de destino

Este capítulo incluye los siguientes temas:

- [Resumen de la optimización de destino, 14](#)
- [Optimización de destino de archivo sin formato, 14](#)
- [Intervalos de puntos de comprobación de base de datos, 15](#)
- [Cargas masivas, 15](#)
- [Optimización de destino de la base de datos, 16](#)

Resumen de la optimización de destino

Optimice los destinos para habilitar el servicio de integración de datos y escribir en los destinos de forma eficiente. Puede soltar índices y limitaciones de clave antes de ejecutar una asignación, aumente el número de intervalos de los puntos de comprobación en la base de datos, configure la carga masiva en las propiedades de escritura para un objeto de datos y optimice una base de datos de destino de Oracle.

Utilice las siguientes técnicas de optimización para optimizar el destino:

- Optimice los destinos de archivo sin formato.
- Aumente los intervalos de punto de comprobación.
- Utilice las cargas masivas.
- Optimice las bases de datos de destino de Oracle.

Optimización de destino de archivo sin formato

Puede mejorar el rendimiento de la asignación optimizando los destinos de archivo sin formato. También puede insertar tareas de transformación en un comando para mejorar el rendimiento.

Tenga en cuenta las siguientes soluciones para reducir los cuellos de botella del destino de archivo sin formato:

Inserte tareas de transformación en un comando en lugar de en el servicio de integración de datos.

Puede mejorar el rendimiento de la asignación insertando tareas de transformación en un comando en lugar de hacerlo en el servicio de integración de datos. También puede utilizar un comando para ordenar o comprimir los datos de destino. En Developer tool, configure la propiedad de comando en las propiedades de tiempo de ejecución para un destino del archivo sin formato.

En UNIX, utilice cualquier comando UNIX válido o secuencia de comandos de shell. En Windows, utilice cualquier comando DOS válido o archivo de procesamiento por lotes. El escritor de archivos sin formato envía datos al comando en lugar de a un destino de archivo sin formato.

Por ejemplo, utilice el siguiente comando para generar un archivo comprimido a partir de los datos de destino:

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

Escriba en un destino de archivo sin formato que sea local para el nodo del proceso de servicio.

Si el servicio de integración de datos se ejecuta en un solo nodo y escribe en un destino de archivo sin formato, puede optimizar el rendimiento de la asignación escribiendo en un destino de archivo sin formato que sea local para el nodo del proceso de servicio.

Intervalos de puntos de comprobación de base de datos

El rendimiento del servicio de integración de datos disminuye cada vez que espera que la base de datos realice un punto de comprobación.

Tenga en cuenta la siguiente solución para reducir los cuellos de botella de los puntos de comprobación de la base de datos:

Aumente el intervalo del punto de comprobación en la base de datos.

Para reducir el número de puntos de comprobación y aumentar el rendimiento, aumente el intervalo de puntos de comprobación en la base de datos.

Aunque aumente el rendimiento cuando reduce el número de puntos de comprobación, también puede aumentar el tiempo de recuperación si la base de datos se cierra inesperadamente.

Cargas masivas

Cuando utiliza la carga masiva, el servicio de integración de datos omite el registro de la base de datos, lo cual acelera el rendimiento.

Tenga en cuenta las siguientes soluciones para reducir los cuellos de botella de la carga masiva:

Configure la carga masiva en las propiedades de escritura de un objeto de datos.

Puede utilizar la carga masiva para mejorar el rendimiento de una asignación que inserta una gran cantidad de datos en una base de datos DB2, ASE Sybase, Oracle o Microsoft SQL Server.

Sin escribir en el registro de la base de datos, la base de datos de destino no puede realizar la reversión. Como resultado, es posible que no pueda realizar la recuperación. Cuando utilice la carga masiva, calcule la importancia del rendimiento mejorado de la asignación frente a la capacidad de recuperar una asignación incompleta.

Optimización de destino de la base de datos

Puede optimizar la base de datos de destino mediante la comprobación de la cláusula de almacenamiento, la asignación de espacio y los segmentos de revertir o deshacer.

Tenga en cuenta las siguientes soluciones para reducir los cuellos de botella de la base de datos de destino:

Compruebe que la base de datos almacena segmentos de revertir o deshacer en los espacios de tablas adecuados, preferiblemente en discos diferentes.

Al escribir en las bases de datos, la base de datos utiliza segmentos de revertir o deshacer durante las cargas. Pida al administrador de la base de datos de que se asegure de que la base de datos almacena segmentos de revertir o deshacer en los espacios de tablas adecuados, preferiblemente en discos diferentes. Los segmentos de reversión o deshacer deberían tener también las cláusulas de almacenamiento adecuadas.

Ajustar el registro de rehacer de la base de datos.

Para optimizar la base de datos, ajuste el registro de rehacer de esta. La base de datos utiliza el registro de rehacer para registrar las operaciones de carga. Asegúrese de que el tamaño del registro rehacer y el tamaño del buffer son óptimos. En una base de datos Oracle, puede ver las propiedades del registro de rehacer en el archivo init.ora.

Conecte con una base de datos Oracle con el protocolo IPC.

Si el servicio de integración de datos se ejecuta en un solo nodo y la instancia de Oracle es local para el nodo del proceso de servicio, puede optimizar el rendimiento mediante el protocolo IPC para conectarse a la base de datos de Oracle. Puede establecer una conexión con la base de datos de Oracle en listener.ora y tnsnames.ora.

CAPÍTULO 3

Optimización de origen

Este capítulo incluye los siguientes temas:

- [Resumen de optimización del origen, 17](#)
- [Optimización de origen de archivo sin formato, 18](#)
- [Optimización de la consulta, 18](#)
- [Filtros condicionales, 19](#)
- [Seleccionar distinto, 19](#)
- [Sugerencias, 20](#)
- [Restricciones, 21](#)
- [Optimización del objeto de datos personalizado, 22](#)
- [Optimización de origen de base de datos, 23](#)

Resumen de optimización del origen

Optimice los orígenes de archivo sin formato, de los objetos de datos relacionales y personalizados para que el servicio de integración de datos pueda leer datos de origen con eficacia.

Utilice las siguientes técnicas de optimización para optimizar los orígenes:

- Lectura de los datos de origen de forma eficaz.
- Uso de las técnicas de optimización de consultas.
- Uso de los filtros condicionales con la consulta SQL.
- Selección de valores únicos desde el origen.
- Aplicar las sugerencias a la consulta SQL.
- Configure las restricciones en objetos de datos lógicos, objetos de datos físicos y tablas virtuales.
- Configurar los objetos de datos personalizados para su optimización.
- Configurar las bases de datos de Oracle, Sybase y Microsoft SQL Server para su optimización.

Optimización de origen de archivo sin formato

Configure las propiedades de formato para orígenes de archivo sin formato para habilitar el servicio de integración de datos y leer los datos de origen eficientemente.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de un origen de archivo sin formato:

No utilice comillas ni caracteres de escape en las propiedades de formato para un archivo sin formato delimitado.

Si se especifica un carácter de escape, el servicio de integración de datos lee el carácter delimitador como un carácter normal incrustado en la cadena. Se puede mejorar ligeramente el rendimiento de la asignación si el archivo de origen no contiene comillas ni caracteres de escape.

Establezca el número de bytes que el servicio de integración de datos lee por línea.

Si la asignación lee desde un origen de archivo sin formato, puede mejorar el rendimiento de la asignación estableciendo el número de bytes que el servicio de integración de datos lee por línea. Configure la propiedad de la longitud del búfer secuencial de línea en las propiedades del tiempo de ejecución para orígenes de archivo sin formato.

De forma predeterminada, el servicio de integración de datos lee 1024 bytes por línea. Si cada línea en el archivo de origen es menor que la configuración predeterminada, puede reducir la longitud del búfer secuencial de línea en las propiedades de la asignación.

Optimización de la consulta

Si una asignación combina varias tablas de origen en un objeto de datos personalizado, puede mejorar el rendimiento optimizando la consulta con las sugerencias de optimización. Asimismo, las instrucciones de selección de tablas sencillas con una cláusula SELECT o GROUP BY pueden beneficiarse de la optimización tal como añadir índices.

Tenga en cuenta las siguientes soluciones para consultar los cuellos de botella:

Cree sugerencias del optimizador para decir a la base de datos cómo ejecutar la consulta para un conjunto determinado de tablas de origen.

Normalmente, el optimizador de la base de datos determina la forma más eficaz para procesar los datos de origen. Sin embargo, es posible que conozca propiedades de las tablas de origen que el optimizador de la base de datos no. El administrador de la base de datos puede crear sugerencias del optimizador para decir a la base de datos cómo ejecutar la consulta para un conjunto determinado de tablas de origen.

Configure las sugerencias del optimizador para comenzar a devolver filas tan rápidamente como sea posible, en lugar de devolver todas las filas de una vez.

Utilice las sugerencias del optimizador si hay un largo retardo desde que la consulta empieza a ejecutarse y el servicio de integración de datos recibe la primera fila de datos. Configure las sugerencias del optimizador para comenzar a devolver filas tan rápidamente como sea posible, en lugar de devolver todas las filas de una vez. Esto permite al servicio de integración de datos procesar filas en paralelo con la ejecución de la consulta.

Cree un índice en las columnas ORDER BY o GROUP BY.

Las consultas que contienen las cláusulas ORDER BY o GROUP BY pueden beneficiarse de crear un índice en las columnas ORDER BY o GROUP BY. Una vez que haya optimizado la consulta, utilice la opción de reemplazo SQL para aprovechar todas las ventajas de estas modificaciones.

Configure la base de datos para ejecutar consultas en paralelo.

También puede configurar la base de datos de origen para ejecutar consultas en paralelo y mejorar el rendimiento. Para obtener más información sobre la configuración de las consultas en paralelo, consulte la documentación de la base de datos.

La consulta que el servicio de integración de datos utiliza para leer datos aparece en la base de datos virtual en un servicio de datos SQL. También puede encontrar la consulta en el objeto de datos personalizado. Pida al administrador de la base de datos que analice la consulta y, a continuación, cree sugerencias del optimizador e índices para las tablas de origen.

Filtros condicionales

Un filtro de origen simple en la base de datos de origen en ocasiones puede afectar al rendimiento negativamente debido a la falta de índices. Puede usar el filtro condicional en el objeto de datos personalizado para mejorar el rendimiento.

Tenga en cuenta la siguiente solución para los cuellos de botella de filtros condicionales:

Utilice el filtro condicional para varias asignaciones que se lean desde el mismo origen de forma simultánea.

Si hay varias asignaciones que leen desde el mismo origen de forma simultánea, el filtro condicional puede mejorar el rendimiento.

Sin embargo, algunas asignaciones pueden tener un rendimiento más rápido si filtra los datos de origen en la base de datos de origen. Puede probar la asignación tanto con el filtro de la base de datos como con el filtro condicional para determinar qué método mejora el rendimiento.

Seleccionar distinto

Puede seleccionar valores únicos de orígenes en un objeto de datos personalizado mediante la opción Seleccionar distinto. Cuando use esta opción, el servicio de integración de datos añadirá una declaración SELECT DISTINCT a la consulta SQL predeterminada.

Tenga en cuenta la siguiente solución para los cuellos de botella de seleccionar distinto:

Utilice la opción Seleccionar distinto para filtrar los datos innecesarios anteriores en el flujo de datos.

Utilice la opción Seleccionar distinto para el objeto de datos personalizado si desea que el servicio de integración de datos seleccione valores únicos desde un origen. Utilice la opción Seleccionar distinto para filtrar los datos innecesarios anteriores en el flujo de datos. Esto puede mejorar el rendimiento.

Por ejemplo, puede utilizar la opción Seleccionar distinto para extraer identificadores de cliente únicos de una tabla de totales de ventas. Cuando utilice el objeto de datos personalizado en una asignación, el servicio de integración de datos filtrará los datos innecesarios en una fase temprana del flujo de datos, con lo que aumentará el rendimiento.

Sugerencias

Puede añadir sugerencias a la consulta SQL de origen para pasar instrucciones a un optimizador de base de datos. El optimizador utiliza las sugerencias para elegir un plan de ejecución de las consultas para acceder al origen.

El campo de sugerencias aparece en la vista **Consulta** de una instancia de objeto de datos relacionales o de un objeto de datos personalizado. La base de datos de origen debe ser Oracle, Sybase, IBM DB2 o Microsoft SQL Server. El campo de sugerencias no aparece para otros tipos de bases de datos.

Cuando el servicio de integración de datos genera la consulta de origen, añade las sugerencias de SQL a la consulta de la misma forma en la que las ha introducido en Developer Tool. El servicio de integración de datos no analiza las sugerencias. Cuando ejecuta la asignación que contiene el origen, el registro de la asignación muestra la consulta con las sugerencias en la consulta.

El servicio de integración de datos inserta las sugerencias de SQL en una posición en la consulta según el tipo de base de datos. Consulte la documentación de la base de datos para obtener información sobre la sintaxis para las sugerencias.

Oracle

El servicio de integración de datos añade sugerencias directamente después de la palabra clave SELECCIONAR/ACTUALIZAR/INSERTAR/SUPRIMIR.

```
SELECCIONAR /*+ <sugerencias> */ DESDE ...
```

El signo '+' indica el inicio de las sugerencias.

Las sugerencias se encuentran en un comentario (/...*/ o--... hasta el final de línea)

Sybase

El servicio de integración de datos añade las sugerencias después de la consulta. Configure un nombre de plan en la sugerencia.

```
SELECCIONAR ... PLAN <plan>
```

```
seleccione avg (precio) desde el plan de títulos "(scalar_agg (i_scan type_price_ix titles )"
```

IBM DB2

Puede introducir la cláusula optimizar para como sugerencia. El servicio de integración de datos añade la cláusula al final de la consulta.

```
SELECCIONAR ... OPTIMIZAR PARA <n> FILAS
```

La cláusula optimizar para le dice al optimizador de la base de datos cuántas filas de la consulta puede procesar. La cláusula no limita el número de filas. Si la base de datos procesa más de <n> filas, se puede reducir el rendimiento.

Microsoft SQL Server

El servicio de integración de datos añade sugerencias al final de la consulta como parte de una cláusula OPTION.

```
SELECCIONAR ... OPCIÓN ( <query_hints> )
```

Reglas y directrices de sugerencias

Utilice las siguientes reglas y directrices cuando configure las sugerencias para las consultas SQL:

- Si habilita la optimización de inserción o si utiliza un semi-join en un objeto de datos relacionales, entonces la consulta de origen original cambia. El servicio de integración de datos no aplica las sugerencias a la consulta modificada.
- Puede combinar sugerencias con reemplazos de combinación y filtro, pero si configura un reemplazo de SQL, este tendrá preferencia y el servicio de integración de datos no aplica los demás reemplazos.
- La vista **Consulta** muestra una vista simple o una vista avanzada. Si especifica una sugerencia con un reemplazo de filtro, orden o combinación en la vista simple, Developer Tool muestra el reemplazo completo de la consulta en la vista avanzada.

Cómo crear sugerencias

Cree sugerencias para enviar instrucciones al optimizador de la base de datos para determinar un plan de consulta.

1. Abra el objeto de datos personalizado o la instancia del objeto de datos relacional.
2. Seleccione la vista **Lectura**.
3. Seleccione la transformación de salida.
4. Seleccione las propiedades de **Consulta**.
5. Seleccione la consulta simple.
6. Haga clic en **Editar**, junto al campo **Sugerencias**.
Aparecerá el cuadro de diálogo **Sugerencias**.
7. Especifique la sugerencia en el campo **Consulta SQL**.
Developer Tool no valida la sugerencia.
8. Haga clic en **Aceptar**.
9. Guarde el objeto de datos.

Restricciones

El servicio de integración de datos puede leer restricciones de orígenes relacionales, orígenes de archivo sin formato, objetos de datos lógicos o tablas virtuales. Una restricción es una expresión condicional que deben cumplir los valores de una fila de datos.

Cuando el servicio de integración de datos lee restricciones, puede quitar las filas que no tengan como resultado TRUE para las filas de datos según el método de optimización aplicado.

Antes de establecer una restricción, debe comprobar que los datos de origen cumplen la condición definida por la restricción.

Por ejemplo, una base de datos de origen tiene una columna EDAD que tiene filas con EDAD < 70. Puede establecer una restricción con EDAD < 70 en la base de datos de origen. El servicio de integración de datos lee registros de la base de datos de origen con la restricción EDAD < 70. Si el servicio de integración de datos lee registros con EDAD >= 70, puede quitar las filas con EDAD >= 70.

En la base de datos, puede usar comandos SQL para definir las restricciones en el entorno de la base de datos cuando se conecta a la base de datos. El servicio de integración de datos ejecuta el SQL de entorno de conexión cada vez que se conecta a la base de datos.

Utilice Developer tool para definir las restricciones en objetos de datos lógicos, objetos de datos físicos y tablas virtuales. Cuando define una restricción, debe especificar una expresión que dé como resultado TRUE para cada fila de datos.

Configurar restricciones

Puede agregar restricciones a los objetos de datos relacionales, objetos de datos de archivos sin formato, objetos de datos personalizados, objetos de datos lógicos y tablas virtuales. Después de añadir una restricción, puede editar o eliminar la restricción.

1. En la vista **Explorador de objetos**, abra la asignación que contiene el objeto de datos relacionales agregado como transformación de lectura. O bien, abra el objeto de datos de archivo sin formato, el objeto de datos personalizados, el objeto de datos lógicos o la tabla virtual.
 - Para establecer restricciones en un objeto de datos relacional agregado a una asignación como transformación de lectura, seleccione la transformación de lectura en la asignación. En la vista **Propiedades**, seleccione la ficha **Avanzada**.
 - Para establecer restricciones en un objeto de datos de archivo sin formato, seleccione la vista **Avanzada** y expanda la sección **Tiempo de ejecución: lectura**.
 - Para establecer restricciones en un objeto de datos personalizado, seleccione la vista **Lectura** y, a continuación, seleccione el puerto de **salida** de la transformación de origen. En la vista **Propiedades**, seleccione la ficha **Avanzada**.
 - Para establecer restricciones en un objeto de datos lógicos, seleccione un modelo de datos lógicos y, a continuación, seleccione el objeto de datos lógicos. En la vista **Propiedades**, seleccione la ficha **Avanzada**.
 - Para establecer restricciones en una tabla virtual, abra la tabla virtual desde el punto final de SQL. En la vista **Propiedades**, seleccione la ficha **Avanzada**.
2. Haga clic en el campo de valores para las restricciones.
Aparecerá el cuadro de diálogo **Restricciones**.
3. Haga clic en **Nuevo** para abrir el editor de expresiones.
4. Configure la restricción lógica y utilice funciones y columnas de expresión como parámetros.
5. Haga clic en **Validar**.
6. Haga clic en **Aceptar**.

Optimización del objeto de datos personalizado

Puede configurar objetos de datos personalizados para mejorar el rendimiento. Puede optimizar la consulta SQL, utilice los filtros condicionales y seleccione valores distintos desde el origen en un objeto de datos personalizado.

Tenga en cuenta las siguientes soluciones para los cuellos de botella del objeto de datos personalizado:
Cree una consulta personalizada para emitir una instrucción SELECT especial para que el servicio de integración de datos lea datos de origen.

La consulta personalizada reemplaza la consulta predeterminada que utiliza el servicio de integración de datos para leer los datos desde los orígenes.

Filtre filas cuando el servicio de integración de datos lea datos de origen.

Si incluye una condición de filtro, el servicio de integración de datos añade una cláusula WHERE a la consulta predeterminada.

Seleccionar valores distintos del origen.

Cuando se usa Seleccionar distinto, el servicio de integración de datos añade una declaración SELECT DISTINCT a la consulta SQL predeterminada.

Aplicar sugerencias de base de datos.

Puede añadir sugerencias a la consulta SQL de origen para pasar instrucciones a un optimizador de base de datos.

Configure las restricciones en los datos de origen.

Si configura las restricciones en archivos sin formato y tablas relacionales de un objeto de datos personalizado, el servicio de integración de datos anulará las filas que no tengan como resultado TRUE para las filas de datos.

Optimización de origen de base de datos

Si la base de datos de origen es de Oracle, puede optimizar el rendimiento del servicio de integración de datos, mediante el protocolo IPC para conectarse a la base de datos Oracle. También puede mover la base de datos temporal a una matriz de disco para mejorar el rendimiento.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de origen de la base de datos:

Utilice el protocolo IPC para conectarse a la base de datos de Oracle.

Si el servicio de integración de datos se ejecuta en un solo nodo y la instancia de Oracle es local para el nodo del proceso de servicio, puede optimizar el rendimiento mediante el protocolo IPC para conectarse a la base de datos de Oracle. Puede establecer una conexión con la base de datos de Oracle en listener.ora y tnsnames.ora.

Mueva la base de datos temporal y los registros de rehacer a una matriz de disco o a unidades más rápidas.

Cuando se combinan tablas de gran tamaño en una base de datos, puede utilizar una matriz redundante de discos independientes (RAID) para la ubicación de la memoria caché. También puede añadir más archivos al grupo de archivos principal de otros discos a fin de dividir la carga entre el discos.

CAPÍTULO 4

Optimización de transformación

Este capítulo incluye los siguientes temas:

- [Optimización de transformación, 24](#)
- [Optimización de la transformación de agregación, 24](#)
- [Optimización de la expresión, 25](#)
- [Optimización de transformación Java, 27](#)
- [Optimización de la transformación de combinación, 30](#)
- [Optimización de la transformación de búsqueda, 30](#)
- [Optimización de la transformación de ordenación, 33](#)
- [Optimización de la transformación de SQL, 34](#)
- [Memoria caché de la transformación, 35](#)
- [Eliminación del error de transformación, 36](#)
- [Efectos secundarios de la transformación, 37](#)
- [Optimización de la transformación del consumidor de servicio web, 37](#)

Optimización de transformación

Optimice las transformaciones para permitir que el servicio de integración de datos procese las transformaciones de una asignación eficientemente.

Utilice las siguientes técnicas de optimización para optimizar la transformación:

- Configure las transformaciones para la optimización.
- Elimine los errores de transformación.
- Configure la memoria caché de la transformación.

Optimización de la transformación de agregación

Las transformaciones de agregación suelen reducir el rendimiento porque deben agrupar datos antes de procesarlos. Las transformaciones de agregación necesitan memoria adicional para albergar los resultados intermedios del grupo.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de la transformación de agregación:

Agrupe por columnas simples.

Puede optimizar las transformaciones de agregación al agrupar por columnas simples. Cuando sea posible, utilice números en lugar de cadenas y fechas en las columnas utilizadas para GROUP BY. Evite expresiones complejas en las expresiones de agregación.

Utilice la entrada ordenada.

Para aumentar el rendimiento de la asignación, ordene los datos para la transformación de agregación. Utilice la opción Entrada ordenada para ordenar los datos.

La opción Entrada ordenada reduce el uso de memorias cachés de agregado. Cuando utiliza la opción Entrada ordenada, el servicio de integración de datos asume que todos los datos están ordenados por grupo. Cuando el servicio de integración de datos lee las filas de un grupo, realiza cálculos de agregado. Si es necesario, almacena la información de grupo en la memoria.

La opción Entrada ordenada reduce la cantidad de datos que se guardan en la memoria caché durante la asignación y mejora el rendimiento. Utilice la opción Entrada ordenada o una transformación de ordenación para pasar los datos ordenados a la transformación de agregación.

Puede aumentar el rendimiento cuando utilice la opción Entrada ordenada en asignaciones con varias particiones.

Filtre los datos antes de agregarlos.

Si utiliza una transformación de filtro en la asignación, colóquela antes de la transformación de agregación para reducir toda agregación innecesaria.

Limite las conexiones a puerto.

Limite el número de puertos de entrada/salida o de salida conectados para reducir la cantidad de datos que la transformación de agregación almacena en la memoria caché de datos.

Optimización de la expresión

Algunas expresiones utilizadas en una transformación podrían reducir el rendimiento.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de la expresión:

Aísle las expresiones lentas.

Las expresiones lentas reducen el rendimiento de la asignación. Para aislar las expresiones lentas, quite las expresiones de la asignación de una en una y ejecute la asignación para determinar el tiempo que se tarda en ejecutar la asignación sin la expresión. Si hay una diferencia significativa en el tiempo de ejecución de la asignación, busque formas para optimizar la expresión lenta.

Realice los siguientes pasos para evaluar el rendimiento de la expresión:

1. Calcule el tiempo de la asignación con las expresiones originales.
2. Copie la asignación y reemplace la mitad de las expresiones complejas con una constante.
3. Ejecute y calcule el tiempo de la asignación editada.
4. Haga otra copia de la asignación y reemplace la otra mitad de las expresiones complejas con una constante.
5. Ejecute y calcule el tiempo de la asignación editada.

Factor fuera de la lógica común.

Si la asignación realiza la misma tarea en varios lugares, reduzca la cantidad de veces que la asignación realiza la tarea moviendo la tarea antes en la asignación. Por ejemplo, tiene una asignación con cinco

tablas de destino. Cada destino requiere una búsqueda del número de seguridad social. En lugar de realizar la búsqueda cinco veces, coloque la transformación de búsqueda en la asignación antes de que se divida el flujo de datos. A continuación, pase los resultados de la búsqueda a los cinco destinos.

Minimizar las llamadas de la función agregada.

Al escribir expresiones, excluya todas las llamadas de la función agregada como sea posible. Cada vez que se utiliza una llamada de la función agregada, el servicio de integración de datos debe buscar y agrupar los datos. Por ejemplo, en la siguiente expresión, el servicio de integración de datos lee COLUMN_A, busca la suma, después lee COLUMN_B, busca la suma y finalmente busca la suma de las dos sumas:

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

Si excluye la llamada de la función agregada, como a continuación, el servicio de integración de datos añade COLUMN_A a COLUMN_B y entonces busca la suma de ambas.

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

Reemplace expresiones comunes con variables locales.

Si utiliza la misma expresión varias veces en una transformación, puede hacer que la expresión sea una variable local. Solo puede utilizar una variable local en la transformación. Sin embargo, al calcular la variable sólo una vez, se aumenta la velocidad del rendimiento.

Seleccione operadores numéricos frente a operadores de cadena.

El servicio de integración de datos procesa las operaciones numéricas más rápido que las operaciones de cadena. Por ejemplo, si busca grandes cantidades de datos en dos columnas EMPLOYEE_NAME y EMPLOYEE_ID, configurar la búsqueda alrededor de EMPLOYEE_ID aumenta el rendimiento.

Optimizar las comparaciones CHAR-CHAR y CHAR-VARCHAR.

Cuando el servicio de integración de datos realiza comparaciones entre columnas CHAR y VARCHAR, cada vez que busca espacios en blanco finales en la fila es más lento. Puede utilizar la opción TreatCHARasCHARonRead cuando configure el servicio de integración de datos en Informatica Administrator para que el servicio de integración de datos no recorte los espacios finales desde el final de los campos de origen del carácter.

Seleccionar DECODE frente a LOOKUP.

Cuando utiliza una función LOOKUP, el servicio de integración de datos debe buscar una tabla en una base de datos. Cuando utiliza una función DECODE, puede incorporar los valores de la búsqueda a la expresión para que el servicio de integración de datos no tenga que buscar otra tabla. Por tanto, si desea buscar un pequeño conjunto de valores que no cambian, utilice DECODE para mejorar el rendimiento.

Utilice operadores en lugar de funciones.

El servicio de integración de datos lee expresiones escritas con operadores más rápidamente que expresiones con funciones. Si es posible, utilice operadores para escribir expresiones. Por ejemplo, tiene la siguiente expresión que contiene funciones CONCAT anidadas:

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

Puede reescribir esa expresión con el || operador de este modo:

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

Optimizar las funciones IIF.

Las funciones IIF pueden devolver un valor y una acción, lo cual permite obtener más expresiones compactas. Por ejemplo, tiene un origen con tres indicadores S/N: FLG_A, FLG_B, FLG_C. Desea devolver los valores según el valor de cada indicador.

Utilice la siguiente expresión:

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

Optimización de transformación Java

Algunas transformaciones de Java en una asignación pueden reducir el rendimiento.

Tenga en cuenta la siguiente solución para aumentar el rendimiento de la transformación de Java:

Habilite la primera selección o los métodos de optimización de filtro de inserción con la transformación de Java.

Puede habilitar la optimización de primera selección o de inserción en las transformaciones de Java.

Actualice los fragmentos de código en la ficha **Interfaces del optimizador** de la transformación de Java.

Optimización de primera selección con la transformación de Java

Puede habilitar una transformación de Java activa o pasiva para la optimización de primera selección si la transformación Java no tiene ningún efecto secundario. El optimizador pasa la lógica de filtro a través de la transformación de Java y modifica la condición de filtro según sea necesario.

Para ver los fragmentos de código para la optimización de primera selección, seleccione

PredicatePushOptimization en el navegador de la ficha **Interfaces del optimizador**.

allowPredicatePush

Booleano. Habilita la primera selección. Cambie la función para devolver un resultado TRUE y el mensaje para poder habilitar la primera selección. El valor predeterminado es FALSE y la función devuelve un mensaje que no es compatible con la optimización.

```
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

Booleano. Indica si la transformación de Java puede devolver o no un error de campo de salida, tal como un error de dividir por cero. Cambie la función devolver un resultado FALSE si la transformación de Java no

genera errores del campo de salida. Cuando la transformación de Java puede generar errores de campos, el servicio de integración de datos no puede utilizar la optimización de primera selección.

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

getInputExpr

Devuelve una expresión de Informatica que describe qué valores de entrada de los campos de entrada tienen un campo de salida. El optimizador necesita saber qué campos de entrada tienen un campo de salida para insertar la lógica de filtro a la transformación.

```
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;
}
```

Por ejemplo, una asignación contiene una expresión de filtro, "out0 > 8". Out0 es el valor del puerto de salida de out0 en la transformación de Java. Puede definir el valor de out0 como el valor del puerto de entrada in0 + 5. El optimizador puede insertar la siguiente expresión "(in0 + 5) > 8" más allá de la transformación de Java con la optimización de primera selección. Puede devolver NULL si un campo de salida no tiene una expresión de campo de entrada. El optimizador no inserta expresiones de filtro después de los campos de salida sin expresiones de entrada.

Puede incluir el siguiente código:

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

inputGroupsPushPredicateTo

Devuelve una lista de grupos que pueden recibir la lógica de filtro. La transformación de Java tiene un grupo de entrada. No modifique esta función para la transformación de Java.

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

```
        return null;
    }
}
```

Optimización de inserción con la transformación de Java

Puede habilitar una transformación de Java activa para la optimización de inserción si no tiene ningún efecto secundario y la optimización no afecta a los resultados de la asignación.

Cuando configure la optimización de inserción para la transformación de Java, puede definir una forma para que la transformación de Java almacene la condición de filtro que recibe desde el optimizador. Añada el código que examina la condición de filtro. Si la transformación de Java puede absorber la lógica de filtro, entonces la transformación de Java devuelve una condición TRUE al optimizador. El optimizador elimina la transformación de filtro de la asignación optimizada.

Cuando configure la transformación de Java escriba el código que almacena la condición de filtro como metadatos de transformación durante la optimización. También puede escribir el código para recuperar la condición de filtro en tiempo de ejecución y para quitar las filas según la lógica de filtro.

Cuando defina la transformación de Java, puede añadir código para la optimización de inserción en la ficha **Interfaces del optimizador** de la transformación de Java. Para acceder a los fragmentos de código para la optimización de inserción, seleccione FilterPushdownOptimization en el navegador de la ficha **Interfaces del optimizador** de la transformación.

Developer Tool muestra los fragmentos de código para habilitar la optimización de inserción y para recibir la condición de filtro del optimizador. Actualice los fragmentos de código para habilitar la optimización y para guardar la lógica de filtro como metadatos de transformación.

isFilterSupported

Devuelve TRUE para habilitar la optimización de inserción. Devuelve FALSE para deshabilitar la optimización de inserción.

Cambie la función para que devuelva TRUE para poder habilitar la optimización de inserción.

```
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

Recibe la condición de filtro del optimizador.

Añada el código para examinar el filtro y determinar si la lógica de filtro puede utilizarse en la transformación. Si la transformación puede absorber el filtro, entonces use el siguiente método para almacenar la condición de filtro como metadatos de transformación:

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

La clave es un identificador de los metadatos. Puede definir cualquier cadena como clave. Los datos son los datos que desea almacenar en orden para determinar qué filas soltar en tiempo de ejecución. Por ejemplo, los datos pueden ser la condición de filtro que la transformación de Java recibe desde el optimizador.

```
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");
}
```

Optimización de la transformación de combinación

Las transformaciones de combinación pueden reducir el rendimiento porque necesitan espacio adicional en tiempo de ejecución para guardar los resultados intermedios.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de la transformación de combinación:

Designe el origen principal como origen con menos valores de clave duplicadas.

Cuando el servicio de integración de datos procesa una transformación de combinación ordenada, guarda simultáneamente en la memoria caché cien claves únicas. Si el origen principal contiene muchas filas con el mismo valor de clave, el servicio de integración de datos debe guardar más filas en la memoria caché, con lo cual se reduce el rendimiento.

Designe el origen principal como origen con menos filas.

La transformación de combinación compara cada fila del origen de detalle con el origen principal. A menor número de filas que haya en el origen principal se producen menos iteraciones de la comparación de uniones, lo que acelera el proceso de unión.

Realice uniones en una base de datos siempre que sea posible.

Las uniones en bases de datos se realizan más rápido que las que se realizan durante la ejecución de la asignación. El rendimiento puede variar en función del tipo de unión de base de datos que se utilice. Las uniones normales son más rápidas que los outer joins y producen menos filas. En ocasiones no es posible realizar la unión en la base de datos, como unir tablas de dos bases de datos o sistemas de archivos sin formato diferentes.

Una datos ordenados siempre que sea posible.

Configure la transformación de combinación para que utilice la entrada ordenada. El servicio de integración de datos aumenta el rendimiento reduciendo al mínimo la entrada y la salida de disco. Se observa el mayor rendimiento al trabajar con conjuntos de datos de gran tamaño. Para una transformación de combinación no ordenada, designe como origen principal el origen con menos filas.

Optimizar la condición de combinación.

El servicio de integración de datos intenta reducir el tamaño del conjunto de datos de un operando de combinación leyendo las filas del grupo más pequeño, buscando las filas coincidentes en el grupo más grande y luego llevando a cabo la operación de combinación. La reducción del tamaño del conjunto de datos mejora el rendimiento de la asignación, ya que el servicio de integración de datos ya no necesita leer filas innecesarias desde el origen del grupo más grande. El servicio de integración de datos mueve la condición de combinación al origen del grupo más grande y solamente lee las filas que coinciden con el grupo más pequeño.

Utilice el método de optimización semi-join.

Utilice el método de optimización semi-join para mejorar el rendimiento de la asignación cuando un grupo de entrada tiene muchas más filas que el otro y cuando el grupo más grande tiene muchas filas sin coincidencia en el grupo más pequeño en función de la condición de combinación.

Optimización de la transformación de búsqueda

Las transformaciones de búsqueda pueden reducir el rendimiento según el tipo de la memoria caché de búsqueda y de las condiciones de búsqueda.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de la transformación de búsqueda:

Utilice el controlador de base de datos óptimo.

El Servicio de integración de datos puede conectarse a una tabla de búsqueda mediante un controlador de base de datos nativo o un controlador ODBC. Los controladores de la base de datos nativos proporcionan un mayor rendimiento de la asignación que los controladores ODBC.

Almacene en la memoria caché las tablas de búsqueda para las búsquedas de archivos relacionales o sin formato.

Para mejorar el rendimiento de la búsqueda de orígenes de archivos relacionales o sin formato, habilite la memoria caché de búsqueda en la transformación. Cuando se habilita la memoria caché, el Servicio de integración de datos almacena en la memoria caché la tabla de búsqueda. Cuando se ejecuta la asignación, el Servicio de integración de datos consulta la memoria caché de búsqueda en lugar de la tabla de búsqueda. Si esta opción no está habilitada, el Servicio de integración de datos consulta la tabla de búsqueda fila por fila.

El resultado de la consulta y el procesamiento de la búsqueda es el mismo, tanto si almacena la tabla de búsqueda en la memoria caché como si no. Sin embargo, al utilizar una memoria caché de búsqueda puede aumentar el rendimiento de la asignación para tablas de búsqueda más pequeñas. En general, querrá almacenar en la memoria caché tablas de búsqueda que necesiten menos de 300 MB.

Almacene en la memoria caché las tablas de búsqueda para las búsquedas de objetos de datos lógicos.

Para mejorar el rendimiento de la búsqueda en un objeto de datos lógicos, puede habilitar la memoria caché de objetos de datos en el Servicio de integración de datos. Cuando se habilita la memoria caché de objetos de datos, el Servicio de integración de datos almacena en la memoria caché el objeto de datos lógicos. Para habilitar la memoria caché de objetos de datos, debe implementar la asignación en una aplicación, habilitar la memoria caché del objeto de datos lógicos y ejecutar la asignación con el comando `infacmd ms runmapping`. Cuando se ejecuta la asignación, el Servicio de integración de datos consulta la memoria caché del objeto de datos en lugar del objeto de datos lógicos.

Si ejecuta la asignación desde Developer tool, la transformación de búsqueda consulta el objeto de datos lógicos fila por fila.

Utilice el tipo de memoria caché correspondiente.

Utilice los siguientes tipos de memoria caché para aumentar el rendimiento:

- Memoria caché compartida. Puede compartir la memoria caché de búsqueda entre varias transformaciones. Puede compartir una memoria caché sin nombre entre transformaciones en la misma asignación. Puede compartir una memoria caché con nombre entre transformaciones en la misma o en diferentes asignaciones.
- Memoria caché persistente. Para guardar y volver a usar los archivos de la memoria caché, puede configurar la transformación para utilizar una memoria caché persistente. Utilice esta característica si sabe que la tabla de búsqueda no cambia entre las ejecuciones de la asignación. Utilizando una memoria caché persistente puede mejorar el rendimiento porque el Servicio de integración de datos compilará la memoria caché desde los archivos de la memoria caché en lugar de desde la base de datos.

Habilitar memorias caché simultáneas.

Cuando el Servicio de integración de datos procesa asignaciones que contienen transformaciones de búsqueda, el Servicio de integración de datos compilará una memoria caché en la memoria cuando procese la primera fila de datos en una transformación de búsqueda almacenada en la memoria caché. Si hay varias transformaciones de búsqueda en una asignación, el Servicio de integración de datos crea las memorias caché secuencialmente cuando la transformación de búsqueda procesa la primera fila de datos. Esto disminuye el procesamiento de la transformación de búsqueda.

Puede habilitar memorias caché simultáneas para mejorar el rendimiento. Cuando el número de canales simultáneos adicionales está establecido en uno o más, el Servicio de integración de datos compilará memorias caché simultáneamente en lugar de secuencialmente. El rendimiento mejora

considerablemente cuando las asignaciones contienen un número de transformaciones activas que puede llevar algún tiempo completar, tales como transformaciones de agregación, combinación u ordenación. Cuando habilite varios canales simultáneos, el Servicio de integración de datos ya no espera a que se completen las asignaciones activas antes de compilar la memoria caché. Otras transformaciones de búsqueda en el canal también construyen memorias caché simultáneamente.

Optimizar coincidencias de condiciones de búsqueda.

Cuando la transformación de búsqueda hace coincidir los datos de la memoria caché de búsqueda con la condición de búsqueda, ordena y clasifica los datos para determinar el primer valor coincidente y el último valor coincidente. Puede configurar la transformación para devolver cualquier valor que coincida con la condición de búsqueda. Cuando configura la transformación de búsqueda para que devuelva cualquier valor coincidente, la transformación devuelve el primer valor que coincide con la condición de búsqueda. No crea un índice de todos los puertos como lo hace cuando configura la transformación para devolver el primer valor coincidente o el último valor coincidente.

Cuando utiliza cualquier valor coincidente, el rendimiento puede mejorar porque la transformación no realiza un índice en todos los puertos, lo cual puede reducir el rendimiento.

Reduzca el número de filas almacenadas en la memoria caché.

Puede reducir el número de filas incluidas en la memoria caché para aumentar el rendimiento. Utilice la opción de reemplazo de SQL de búsqueda para añadir una cláusula WHERE a la instrucción SQL predeterminada. Cuando añada una cláusula WHERE a una transformación de búsqueda que usa una caché dinámica, utilice una transformación de filtro antes de la transformación de búsqueda para pasar filas a la caché dinámica que coincidan con la cláusula WHERE.

Reemplazar la instrucción ORDER BY.

De forma predeterminada, el Servicio de integración de datos genera una instrucción ORDER BY para una búsqueda almacenada en la memoria caché. La instrucción ORDER BY contiene todos los puertos de búsqueda. Para aumentar el rendimiento, suprima la instrucción predeterminada ORDER BY y especifique un ORDER BY de reemplazo con menos columnas.

El Servicio de integración de datos siempre genera una instrucción ORDER BY, incluso si especifica una en el reemplazo. Ponga dos guiones '--' tras el reemplazo ORDER BY para suprimir la instrucción ORDER BY generada.

Por ejemplo, una transformación de búsqueda emplea la siguiente condición de búsqueda:

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

La transformación de búsqueda incluye tres puertos de búsqueda utilizados en la asignación, ITEM_ID, ITEM_NAME, y PRICE. Cuando se especifica la instrucción ORDER BY, introduzca las columnas en el mismo orden que los puertos en la condición de búsqueda. También debe delimitar todas las palabras reservadas de la base entre comillas.

Introduzca la siguiente consulta de búsqueda en el reemplazo de SQL de búsqueda:

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

Utilice un equipo con más memoria.

Para aumentar el rendimiento de la asignación, ejecute la asignación en un nodo del Servicio de integración de datos con una gran cantidad de memoria. Aumente el tamaño del índice y de la memoria caché de datos todo lo que sea posible sin forzar el equipo. Si el nodo del Servicio de integración de datos tiene memoria suficiente, aumente la memoria caché de modo que pueda incluir todos los datos en la memoria sin tener que paginar en el disco.

Optimice la condición de búsqueda.

Si incluye más de una condición de búsqueda, coloque las condiciones en el orden siguiente para optimizar el rendimiento de búsqueda:

- Igual a (=)
- Menor que (<), mayor que (>), menor o igual que (<=), mayor o igual que o (>=)
- Distinto de (!=)

Filtre filas de búsqueda.

Para mejorar el rendimiento, cree una condición de filtro para reducir el número de filas de búsqueda recuperadas desde el origen cuando la memoria caché de búsqueda se ha construido.

Realice un índice de la tabla de búsqueda.

El Servicio de integración de datos necesita consultar, ordenar y comparar valores en las columnas de la condición de búsqueda. El índice debe incluir cada columna utilizada en una condición de búsqueda.

Puede mejorar el rendimiento de los siguientes tipos de búsquedas:

- Búsquedas almacenadas en la memoria caché. Para mejorar el rendimiento, realice un índice de las columnas en la instrucción de búsqueda ORDER BY. El archivo de registro de la asignación contiene la instrucción ORDER BY.
- Búsquedas no almacenadas en la memoria caché. Para mejorar el rendimiento, realice un índice de las columnas en la condición de búsqueda. El Servicio de integración de datos emite una instrucción SELECT para cada fila que pasa por la transformación de búsqueda.

Optimice varias búsquedas.

Si una asignación contiene varias búsquedas, incluso con la memoria caché habilitada y suficiente memoria de heap, las búsquedas pueden reducir el rendimiento. Ajuste las transformaciones de búsqueda que consultan las cantidades de datos más grandes para mejorar el rendimiento general.

Si la tabla de búsqueda está en la misma base de datos que la tabla de origen en la asignación y la memoria caché no es viable, combine las tablas en la base de datos de origen en lugar de utilizar una transformación de búsqueda.

Optimización de la transformación de ordenación

Las transformaciones de ordenación pueden reducir el rendimiento cuando la RAM física en el nodo del servicio de integración de datos no tiene suficiente memoria asignada para ordenar los datos.

Tenga en cuenta la siguiente solución para los cuellos de botella de la transformación de ordenación:

Asigne suficiente memoria.

Para un rendimiento óptimo, configure el tamaño de la memoria caché de ordenación con un valor menor o igual que la cantidad de memoria física RAM del nodo del servicio de integración de datos. Asigne como mínimo 16 MB de memoria física para ordenar datos con la transformación de ordenación. De forma predeterminada, el tamaño de la memoria caché de ordenación se establece en 16.777.216 bytes. Si el servicio de integración de datos no puede asignar suficiente memoria para ordenar los datos, la asignación falla.

Si la cantidad de datos entrantes es mayor que la cantidad de memoria caché de ordenación, el servicio de integración de datos almacena datos temporalmente en el directorio de trabajo de la transformación de ordenación. El servicio de integración de datos necesita un espacio de disco de como mínimo el doble de la cantidad de datos entrantes cuando ordena datos en el directorio de trabajo.

Optimización de la transformación de SQL

Cada vez que el servicio de integración de datos procesa una nueva consulta en una asignación, llama a una función denominada SQLPrepare para crear un procedimiento de SQL y pasarlo a la base de datos. Cuando la consulta cambia para cada fila de entrada, se podría reducir el rendimiento.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de la transformación de SQL:

No utilice instrucciones de transacción en una consulta de transformación de SQL.

Cuando una consulta SQL contiene instrucciones de consulta de confirmación y reversión, el servicio de integración de datos debe recrear el procedimiento de SQL después de cada confirmación o reversión. Para optimizar el rendimiento, no use instrucciones de transacción en una consulta de transformación de SQL.

Habilite los métodos de optimización de filtro de primera selección o de inserción con la transformación SQL.

Para aumentar el rendimiento, podría habilitar el método de optimización de primera selección o de inserción con la transformación de SQL.

Optimización de primera selección con la transformación de SQL

El servicio de integración de datos puede realizar la optimización de primera selección con una transformación de SQL si la condición de filtro hace referencia solo a los puertos de transferencia y la transformación de SQL no tiene efectos secundarios.

La transformación de SQL tiene efectos secundarios en las siguientes circunstancias:

- La consulta SQL actualiza una base de datos. La consulta SQL contiene una instrucción como CREATE, DROP, INSERT, UPDATE, GRANT o REVOKE.
- La transformación devuelve filas nulas para las instrucciones SELECT que no devuelven ningún resultado. Las filas pueden contener valores de puerto de transferencia, información de errores de SQL o el campo NumRowsAffected.

Habilitar la optimización de primera selección con la transformación de SQL

Habilite la optimización de primera selección en la transformación de SQL si la transformación de SQL no tiene ningún efecto secundario.

1. Habilite la opción **Devolver la base de datos de solo salida** en **Propiedades avanzadas** de la transformación de SQL.
2. Desactive **Tiene efectos secundarios** en **Propiedades avanzadas** de la transformación.
3. Si la transformación tiene un puerto **NumAffectedRows**, quite el puerto.

Optimización de inserción con la transformación de SQL

Con la optimización de inserción, el servicio de integración de datos inserta la lógica de filtro desde una transformación de filtro de la asignación a la consulta de la transformación de SQL.

Utilice las siguientes reglas y directrices cuando habilite la optimización de inserción con la transformación de SQL:

- La transformación de la consulta SQL solo debe contener instrucciones SELECT.
- La consulta SQL de la transformación debe ser una subconsulta válida.
- La condición de filtro no puede hacer referencia a los campos de Error de SQL o NumRowsAffected.

- Los nombres de los puertos de salida deben coincidir con los nombres de las columnas de la instrucción SQL SELECT. Cuando hace referencia a un puerto de salida en una condición de filtro, el servicio de integración de datos inserta el nombre del puerto correspondiente en la consulta SQL. Puede agregar alias a SQL si las columnas de la consulta no coinciden con los nombres de puertos de salida. Por ejemplo, `SELECT mycolname1 AS portname1, mycolname2 AS portname2`.
- La transformación no puede tener efectos secundarios.

Ejemplo de la optimización de inserción con la transformación SQL

Una transformación de SQL recupera pedidos por el ID del cliente. Una transformación de filtro que aparece tras la transformación de SQL solamente devuelve las filas en las que el importe del pedido es superior a 1.000.

El servicio de integración de datos inserta el siguiente filtro en una instrucción SELECT en la transformación de SQL:

```
orderAmount > 1000
```

Cada instrucción de la consulta SQL se convierte en otra subconsulta de la instrucción SELECT que contiene el filtro.

La siguiente instrucción de consulta muestra la instrucción de consulta original como una subconsulta de la instrucción SELECT:

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

Si la consulta SQL tiene varias instrucciones, cada instrucción está incluida en otra subconsulta. La subconsulta tiene la misma sintaxis, incluyendo la cláusula WHERE.

Los puertos *customerID* y *orderAmount* son los nombres de los puertos de salida en la transformación de SQL. La subconsulta no incluye puertos de transferencia, el error de SQL o los puertos de estadísticas de SQL. Si inserta varios filtros en la transformación de SQL, la cláusula WHERE contiene todos los filtros.

Habilitar la optimización de inserción con la transformación de SQL

Habilitar la optimización de inserción mediante las propiedades de configuración en la ficha **Propiedades avanzadas** de la transformación de SQL.

1. Desactive **Tiene efectos secundarios**.
2. Habilite **Devolver la base de datos de solo salida**.
3. Establezca el **Recuento máximo de filas de salida** a cero.
4. Habilitar la optimización de inserción.

Memoria caché de la transformación

Cuando se ejecuta una asignación que utiliza una transformación de agregación, combinación, búsqueda, rango u ordenación, el Servicio de integración de datos crea memorias caché en la memoria para procesar la transformación. Si el Servicio de integración de datos requiere más espacio, almacena los valores desbordados en archivos de memoria caché en el disco.

Tenga en cuenta la siguiente solución para los cuellos de botella de la memoria caché de la transformación:

Configure las transformaciones para asignar espacio suficiente para almacenar la memoria caché en la memoria.

Para reducir el tiempo de procesamiento de las transformaciones de agregación, combinación, búsqueda, rango u ordenación, configure las transformaciones para asignar suficiente espacio para almacenar la memoria caché en la memoria. Cuando configura la cantidad de memoria caché para que sea igual o mayor que lo que se requiere para almacenar los datos y el índice en la memoria caché, aumenta el rendimiento reduciendo la sobrecarga de E/S del sistema. Cuando el Servicio de integración de datos escribe archivos de memoria caché en el disco, el tiempo de procesamiento aumenta debido a la sobrecarga de E/S del sistema.

De forma predeterminada, el Servicio de integración de datos configura automáticamente los requisitos de la memoria caché en tiempo de ejecución. Después de ejecutar una asignación en modo de memoria caché automática, puede ajustar los tamaños de la memoria caché para las transformaciones. Analice las estadísticas de transformación en el registro de la asignación para determinar los tamaños de memoria caché necesarios para procesar las transformaciones en la memoria. Al configurar el tamaño de memoria caché para que use el valor especificado en el registro de la asignación, se asegurará de que no se desperdicie nada de la memoria asignada. Sin embargo, el tamaño de memoria caché óptimo varía según el tamaño de los datos de origen. Revise los registros de la asignación después de ejecutar la asignación posteriormente para supervisar los cambios en el tamaño de la memoria caché. Si configura un tamaño de memoria caché específico para una transformación reutilizable, compruebe que el tamaño de la memoria caché sea óptimo para cada uso de la transformación en una asignación.

Eliminación del error de transformación

En números grandes, los errores de transformación reducen el rendimiento del servicio de integración de datos. Con cada error de transformación, el servicio de integración de datos se detiene para determinar la causa del error y para eliminar la fila que está provocando el error desde el flujo de datos. El servicio de integración de datos normalmente escribe la fila en el archivo de registro de asignación en los registros del servicio de integración de datos.

Tenga en cuenta las siguientes soluciones para los cuellos de botella del error de transformación:

Compruebe el archivo de registro de la asignación para ver dónde se producen los errores de transformación y evalúe las restricciones de transformación.

Los errores de transformación se producen cuando el servicio de integración de datos detecta errores de conversión, lógica de asignación en conflicto y cualquier condición configurada como un error, tales como entrada nula. Compruebe el archivo de registro de la asignación para ver dónde se producen los errores de transformación. Si los errores se centran alrededor de determinadas transformaciones, evalúe las restricciones de transformación.

Configure un nivel de seguimiento inferior.

Si necesita ejecutar una asignación que genere un gran número de errores de transformación, es posible mejorar el rendimiento estableciendo un nivel de seguimiento inferior. Sin embargo, esto no es una solución a largo plazo recomendada para los errores de transformación.

Efectos secundarios de la transformación

Una transformación tiene efectos secundarios si devuelve filas y modifica un objeto, o si interactúa con otros objetos o funciones. La transformación puede modificar una base de datos, añadir a un total, elevar una excepción, escribir un correo electrónico o llamar a otras funciones con efectos secundarios.

El servicio de integración de datos identifica qué transformaciones tienen efectos secundarios antes de que optimice una asignación. El servicio de integración de datos asume que una transformación tiene efectos secundarios cuando no se puede determinar si la transformación tiene efectos secundarios.

Las transformaciones con efectos secundarios limitan cuando el servicio de integración de datos puede optimizar la asignación. La optimización de primera selección, de eliminación de ramificaciones, de predicado global y de inserción modifican los resultados de la asignación si el Servicio de integración de datos las aplica a una transformación que tiene efectos secundarios. Con la optimización de primera selección y de inserción, la lógica de filtro se mueve desde una transformación de filtro tan cerca del origen como sea posible. Si el filtro se produce antes de la función de efectos secundarios, los resultados de asignación cambian.

Por ejemplo, una transformación recibe un ID de cliente y devuelve filas que contienen información de los pedidos. La transformación también escribe los pedidos en un archivo. Si el servicio de integración de datos aplica una optimización del filtro antes de que escriba los pedidos en el archivo, el archivo recibe menos filas cuando tiene lugar el filtro más adelante en la asignación. Los efectos secundarios de la transformación es la función de escribir los registros de los pedidos en un archivo.

Las siguientes transformaciones tienen efectos secundarios:

- Transformación de SQL, transformación del consumidor de servicio web, transformación de Java a menos que la propiedad de efectos secundarios esté deshabilitada.
- Las transformaciones que llaman a una función `ABORT()` o `ERROR()`, envían correo electrónico o llaman a un procedimiento almacenado.
- Transformaciones que escriben en archivos o bases de datos.
- Transformaciones que mantienen un recuento mediante un puerto variable. Por ejemplo, `COUNT=COUNT +1`.

La transformación de SQL, la transformación del consumidor de servicio web y la transformación de Java tienen efectos secundarios de forma predeterminada. Si configura la transformación para procesar filas sin efectos secundarios, puede deshabilitar la propiedad **Tiene efectos secundarios** en **Propiedades avanzadas**. Si la transformación no tiene efectos secundarios, puede habilitar la optimización mediante la configuración de las propiedades adicionales en estas transformaciones.

Optimización de la transformación del consumidor de servicio web

La transformación de consumidor de servicio web puede reducir el rendimiento cuando una asignación llama varias veces al servicio web.

Tenga en cuenta la siguiente solución para los cuellos de botella de la transformación del consumidor de servicio web:

Configure la transformación de consumidor de servicio web para que utilice la autenticación por cookies.

El servidor remoto del servicio web hace un seguimiento de los usuarios del servicio web mediante cookies. Es posible aumentar el rendimiento cuando una asignación llama un servicio web varias veces.

Cuando se proyecta el puerto de cookie a un mensaje de solicitud de servicio web, el proveedor del servicio web devuelve un valor de cookie en el mensaje de respuesta. Puede pasar el valor de cookie a otra transformación situada en un nivel inferior en la asignación o puede guardar el valor de la cookie en un archivo. Cuando se guarda el valor de la cookie en un archivo, puede configurar la cookie como entrada de la transformación de consumidor de servicio web. Puede proyectar el puerto de salida de la cookie a cualquiera de los grupos de salida de la transformación de consumidor de servicio web.

Habilite los métodos de optimización de filtro de primera selección o de inserción con la transformación de consumidor de servicio web.

Para aumentar el rendimiento, el servicio de integración de datos puede aplicar el método de optimización de filtro de primera selección o de inserción con la transformación del consumidor de servicio web. Para aplicar la optimización de primera selección, el servicio web no puede tener efectos secundarios y no se puede tratar los fallos como errores. Para aplicar la optimización de inserción, el servicio web no puede tener efectos secundarios, no se puede tratar los fallos como errores y la condición de filtro debe hacer referencia a los puertos de transferencia.

El servicio web tiene efectos secundario si realiza otras funciones además de devolver una respuesta a la transformación del consumidor de servicio web. El servicio web tiene efectos secundarios si modifica una base de datos, escribe en un archivo, escribe un correo electrónico, actualiza un recuento o llama a otros servicios web con efectos secundarios.

Configure la transformación de consumidor del servicio web para enviar varias solicitudes en paralelo.

Para aumentar el rendimiento de la asignación, configure la transformación de consumidor del servicio web para enviar varias solicitudes en paralelo. Cuando habilita la transformación de consumidor del servicio web para crear varias conexiones simultáneas en el servicio web, puede establecer el límite del consumo de memoria y el número de límites de conexiones simultáneas.

Optimización de primera selección con la transformación del consumidor de servicio web

Cuando el servicio de integración de datos aplica el método de optimización de primera selección a la transformación del consumidor de servicio web, mueve las condiciones de filtro antes que la transformación del consumidor de servicio web en la asignación que esté más cerca del origen.

Habilitar la optimización de primera selección con la transformación del consumidor de servicio web

Habilite la optimización de primera selección para la transformación de consumidor de servicio web si la transformación no tiene efectos secundarios y no trata los fallos como errores.

1. Abra la vista **Propiedades avanzadas** de la transformación del consumidor de servicio web.
2. Desactive **Tratar fallo como error**.
3. Desactive **Tiene efectos secundarios**.

Optimización de inserción con la transformación del consumidor de servicio web

Puede configurar la optimización de inserción con la transformación de consumidor de servicio web cuando la transformación está en una tabla virtual en un servicio de datos SQL.

La asignación llama al servicio web para recuperar un conjunto o subconjunto de datos según las instrucciones en la consulta SQL de usuario final. La consulta SQL de usuario final contiene una condición de filtro opcional.

Con la optimización de inserción, la transformación del consumidor de servicio web recibe el valor de filtro en un puerto de filtro. El puerto de filtro es un puerto de entrada inconexo que identifica como un puerto de filtro al configurar la optimización de inserción. El puerto de filtro tiene un valor predeterminado que garantiza que el servicio web devuelve todas las filas si la consulta del usuario final no contiene ningún filtro. El puerto de filtro no es un puerto de transferencia.

Nota: El campo de filtro debe formar parte del grupo raíz en la solicitud de servicio web.

Cuando configure un puerto de filtro, puede identificar un puerto de salida en la transformación de consumidor del servicio web que recibe los datos de la columna desde la respuesta de servicio web. Por ejemplo, si el puerto de filtro es un puerto de entrada llamado EmployeeID, el puerto de salida de la respuesta puede ser un puerto llamado EmployeeNum. Developer Tool debe asociar el puerto de entrada de filtro y un puerto de salida para insertar la lógica de filtro desde la lectura de tabla virtual con la solicitud del consumidor de servicio web. Los puertos de entrada para una solicitud de servicio web suelen ser distintos a los puertos de salida de la respuesta de servicio web.

El campo de filtro no puede ser un puerto de transferencia. Cuando configura un puerto de filtro, el valor predeterminado del puerto cambia al valor de la condición de filtro, por lo que el valor del puerto de salida de transferencia cambia. Un filtro basado en el puerto de transferencia de salida devuelve resultados inesperados.

Puede insertar varias expresiones de filtro para la transformación de consumidor de servicio web. Cada condición de filtro debe tener el siguiente formato:

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

Las condiciones de filtro deben estar unidas por AND. No puede combinar las condiciones con un OR.

Optimización de inserción con el ejemplo de transformación del consumidor de servicio web

Un servicio de datos SQL devuelve pedidos de todos los clientes o devuelve pedidos de un cliente específico según la consulta SQL que reciba del usuario.

El servicio de datos contiene un objeto de datos lógicos con los siguientes componentes:

Tabla de clientes

Una tabla de base de datos Oracle que contiene la información de los clientes.

Transformación de consumidor de servicio web

Una transformación que llama a un servicio web para recuperar los últimos pedidos de los clientes. La transformación del consumidor de servicio web tiene puertos de entrada para customerID y orderNum. La transformación tiene puertos de transferencia que contienen datos de los clientes que recibe desde la tabla de clientes. El puerto orderNum es el puerto de filtro y no está conectado. orderNum tiene el valor predeterminado "*". Cuando el servicio web recibe este valor en la solicitud de servicio web, devuelve todos los pedidos.

Tabla virtual de pedidos

Una tabla virtual que recibe los datos de los clientes y los pedidos desde el servicio web. El usuario final consulta esta tabla. Los pedidos contienen una columna de cliente, una columna orderID y datos de clientes y pedidos.

El usuario final pasa la siguiente consulta SQL al servicio de datos SQL:

```
SELECT * de OrdersID donde cliente = 23 y orderID = 56
```

El servicio de integración de datos divide la consulta para optimizar la asignación. El servicio de integración de datos utiliza la optimización de primera selección y mueve la lógica de filtro, `cliente = 23`, a la lectura de tabla de clientes. El servicio de integración de datos utiliza la optimización de inserción e inserta la lógica de filtro, `orderID = 56`, en el puerto del filtro de la transformación del consumidor de servicio web. La transformación del consumidor de servicio web recupera ordersID 56 para clientes 23.

Habilitar la optimización de inserción con la transformación del consumidor de servicio web

Habilite la optimización de inserción para la transformación de consumidor de servicio web si la transformación no tiene efectos secundarios y no trata los fallos como errores.

1. Abra la vista **Propiedades avanzadas** de la transformación del consumidor de servicio web.
2. Desactive **Tratar fallo como error**.
3. Desactive **Tiene efectos secundarios**.
4. Haga clic en el botón **Abrir** de la propiedad **optimización de inserción**.
5. Seleccione el nombre del puerto de filtro en el cuadro de diálogo Entrada optimizada. Puede seleccionar varios puertos de filtro.
6. Haga clic en la columna **Salida**.
7. Para cada puerto de filtro, seleccione el puerto de salida que contiene la columna filtrada en la respuesta de servicio web.
8. Especifique un valor predeterminado para cada puerto de filtro.

Nota: No puede configurar un valor predeterminado para un puerto del consumidor de servicio web a menos que sea un puerto de filtro.

CAPÍTULO 5

Optimización de la asignación

Este capítulo incluye los siguientes temas:

- [Resumen de optimización de la asignación, 41](#)
- [Métodos de optimización, 42](#)
- [Optimización de inserciones, 49](#)
- [Lectura de pase individual, 52](#)
- [Optimización del filtro, 52](#)
- [Optimización de conversión de tipos de datos, 53](#)
- [Error de seguimiento, 53](#)

Resumen de optimización de la asignación

Optimice las asignaciones para habilitar el servicio de integración de datos para transformar y mover datos eficientemente. La optimización a nivel de asignación puede tardar tiempo en implementarse, pero puede aumentar significativamente el rendimiento de la asignación.

Las tareas de optimización se aplican a asignaciones periódicas, asignaciones de lectura y escritura de objetos de datos lógicos, asignaciones de tablas virtuales y asignaciones operativas. Céntrese en la optimización a nivel de asignación después de optimizar los destinos y orígenes.

Para optimizar una asignación, puede realizar las siguientes tareas:

- Configurar la asignación con el número mínimo de transformaciones y expresiones para realizar la mayor cantidad de trabajo posible.
- Eliminar los vínculos innecesarios entre transformaciones para reducir la cantidad de datos movidos.
- Elegir un nivel del optimizador que determine los métodos de optimización que el servicio de integración de datos puede aplicar a la asignación. Cuando el servicio de integración de datos optimiza una asignación, intenta reducir la cantidad de datos que se procesan. Por ejemplo, el servicio de integración de datos puede utilizar la optimización de primera selección para mover un filtro más cerca del origen. Puede utilizar el método de optimización basado en el coste para cambiar el orden de procesamiento de combinación.
- Elija un tipo de inserción para permitir que el Servicio de integración de datos determine si puede insertar lógica de transformación parcial o completa en la base de datos de origen.
- Configurar la memoria caché de objetos de datos para habilitar los objetos de datos lógicos de la memoria caché en el servicio de integración de datos y acceder a objetos de datos lógicos pregenerados cuando ejecute una asignación. De forma predeterminada, el servicio de integración de datos extrae los datos de origen y genera los objetos de datos requeridos cuando ejecuta una asignación. El rendimiento

de la asignación aumenta cuando el servicio de integración de datos puede acceder a objetos de datos generados previamente.

- Indicar si la transformación de SQL, la transformación del consumidor de servicio web y la transformación de Java no tienen efectos secundarios cuando configure estas transformaciones. Algunas transformaciones tienen efectos secundarios que restringen la optimización. Por ejemplo, una transformación puede tener efectos secundarios si la transformación escribe en un archivo o base de datos, añade a un recuento, genera una excepción o escribe un mensaje de correo electrónico. En la mayoría de los casos, el servicio de integración de datos identifica qué transformaciones tienen efectos secundarios que restringen la optimización.

TEMAS RELACIONADOS

- [“Memoria caché de objetos de datos” en la página 66](#)

Métodos de optimización

El Servicio de integración de datos aplica los métodos de optimización para reducir el número de filas en la asignación. Puede configurar el nivel del optimizador para la asignación para limitar los métodos de optimización a los que se aplica el Servicio de integración de datos.

El Servicio de integración de datos puede aplicar los siguientes métodos de optimización:

- Optimización de inserciones
- Optimización de primera proyección
- Optimización de primera selección
- Optimización de eliminación de ramificaciones
- Optimización de inserción
- Optimización de predicado
- Optimización de predicado global
- Optimización basada en el coste
- Optimización dataship-join
- Optimización semi-join

El Servicio de integración de datos puede aplicar varios métodos de optimización a una asignación al mismo tiempo. Por ejemplo, el Servicio de integración de datos aplica la optimización de primera proyección, la optimización de predicado, la optimización de predicado global, la optimización de eliminación de ramificaciones y la optimización de primera selección o métodos de optimización de inserción cuando seleccione el nivel de optimizador normal.

Niveles del optimizador

El Servicio de integración de datos optimiza las asignaciones en función del nivel del optimizador que configure. Configure la propiedad de nivel de optimizador cuando desee que la asignación use un nivel de optimizador distinto del predeterminado.

Puede elegir uno de los siguientes niveles del optimizador:

Auto

El servicio de integración de datos aplica optimizaciones en función del modo de ejecución y el contenido de la asignación.

Ninguno

El Servicio de integración de Datos no aplica ninguna optimización.

Mínimo

El Servicio de integración de Datos aplica el método de optimización de primera proyección.

Normal

El servicio de integración de datos aplica los métodos de optimización de primera proyección, de primera selección, de eliminación de ramificaciones, de inserción, de predicado global y de predicado.

Completo

El Servicio de integración de datos aplica los métodos de optimización basada en el coste, de primera proyección, de primera selección, de eliminación de ramificaciones, de predicado, de inserción, semi-join y dataship-join.

El valor predeterminado es Auto.

El Servicio de integración de datos aplica el nivel de optimizador normal cuando ejecuta una asignación desde el menú **Ejecutar** o el editor de la asignación de Developer tool. Cuando ejecuta la asignación desde el cuadro de diálogo **Ejecutar**, el Servicio de integración de datos aplica el nivel del optimizador en la configuración de la asignación. Cuando ejecuta la asignación desde la línea de comandos, el Servicio de integración de datos aplica el nivel del optimizador desde las propiedades de la implementación de la asignación en la aplicación.

Nota: El Servicio de integración de datos no aplica el método de optimización de inserciones con un nivel del optimizador. Puede configurar la optimización de inserciones para una asignación en las propiedades de tiempo de ejecución de la asignación.

Optimizaciones de filtro

Las optimizaciones de filtro aumentan el rendimiento reduciendo el número de filas que pasan por la asignación. El servicio de integración de datos puede aplicar la optimización de primera selección o la de inserción.

Cuando el servicio de integración de datos aplica un método de optimización del filtro, mueve un filtro tan cerca del origen como sea posible en una asignación. Si el servicio de integración de datos no puede mover un filtro antes de una transformación en una asignación, es posible que pueda insertar la lógica de filtro en una transformación.

Método de optimización de primera proyección

Cuando el Servicio de integración de datos aplica el método de optimización de primera proyección, identifica los puertos no utilizados y elimina los vínculos entre esos puertos.

El método de optimización de primera proyección mejora el rendimiento reduciendo la cantidad de datos que el Servicio de integración de datos mueve a través de las transformaciones. Cuando el Servicio de integración de datos procesa una asignación, mueve los datos desde todos los puertos conectados en una asignación de una transformación a otra. En asignaciones complejas, de gran tamaño o en asignaciones que utilizan mapplets anidados, es posible que algunos puertos no lleguen a suministrar datos al destino. El Servicio de integración de Datos identifica los puertos que no suministran datos al destino. Después de que

el Servicio de integración de datos haya identificado los puertos no utilizados, elimina de la asignación los vínculos que hay entre todos los puertos no utilizados.

El Servicio de integración de datos no elimina todos los vínculos. Por ejemplo, no quita los siguientes vínculos:

- Vínculos conectados a una transformación que tiene efectos secundarios.
- Vínculos conectados a transformaciones que llaman a una función `ABORT()` o `ERROR()`, que envían correo electrónico o que llaman a un procedimiento almacenado.

Si el Servicio de integración de datos determina que ninguno de los puertos de una transformación se utiliza, elimina todos los vínculos de transformación excepto el vínculo al puerto con el menor número de datos. El Servicio de integración de datos no elimina la transformación no utilizada de la asignación.

Developer Tool habilita este método de optimización de forma predeterminada.

Método de optimización de predicado

Cuando el Servicio de integración de datos aplica el método de optimización de predicado, examina las expresiones de predicado que genera una asignación. Determina si puede simplificar o reescribir las expresiones para aumentar el rendimiento de la asignación.

Cuando el Servicio de integración de datos ejecuta una asignación, genera consultas para los orígenes de asignación y lleva a cabo operaciones en los resultados de la consulta conforme a la lógica de la asignación y a las transformaciones de la asignación. A menudo, las consultas y las operaciones incluyen expresiones de predicado. Las expresiones de predicado representan las condiciones que deben cumplir los datos. Las condiciones de filtro y de unión de las transformaciones de filtro e incorporación son ejemplos de expresiones de predicado.

Con el método de optimización de predicado, el Servicio de integración de datos también intenta aplicar expresiones de predicado tan pronto como sea posible en la asignación para mejorar el rendimiento de la asignación.

El Servicio de integración de datos deduce las relaciones mediante las expresiones de predicado existentes y crea nuevas expresiones de predicado. Por ejemplo, una asignación contiene una transformación de incorporación con la condición de combinación "A=B" y una transformación de filtro con la condición de filtro "A>5". El Servicio de integración de datos puede añadir "B>5" a la condición de combinación.

El Servicio de integración de datos aplica el método de optimización de predicado con el método de optimización de primera selección en los casos en que pueda aplicar ambos métodos a una asignación. Por ejemplo, cuando el Servicio de integración de datos crea nuevas condiciones de filtro mediante el método de optimización de predicado, también intenta moverlas a un nivel superior de la asignación mediante el método de primera selección. Al aplicar ambos métodos de optimización, se mejora el rendimiento de la asignación respecto a la aplicación de sólo uno de los métodos.

El Servicio de integración de datos aplica el método de optimización de predicado si la aplicación aumenta el rendimiento. El Servicio de integración de datos no aplica este método si la aplicación cambia los resultados de la asignación o reduce el rendimiento de la asignación. El Servicio de integración de datos aplica este método de optimización de forma predeterminada.

Reglas y directrices de la optimización de predicado

Cuando el servicio de integración de datos reescribe una expresión de predicado, aplica una lógica matemática a la expresión para optimizarla.

El servicio de integración de datos puede llevar a cabo una de las siguientes acciones, o todas ellas:

- Identifica variables equivalentes en expresiones de predicado de la asignación y genera expresiones simplificadas basándose en las equivalencias.
- Identifica predicados superfluos en expresiones de predicado de la asignación y los elimina.
- Extrae subexpresiones de cláusulas disyuntivas y genera varias expresiones simplificadas basándose en las subexpresiones.
- Normaliza una expresión de predicado.
- Aplica expresiones de predicado lo antes posible en la asignación.

El servicio de integración de datos puede no aplicar una optimización de predicado a una asignación cuando ésta contiene transformaciones con un tipo de datos incongruente entre puertos conectados.

El servicio de integración de datos puede no aplicar la optimización de predicado a una transformación cuando se cumple alguna de las siguientes condiciones:

- La transformación contiene valores predeterminados explícitos para los puertos conectados.
- La transformación tiene efectos secundarios.
- La transformación no permite que se muevan los predicados. Por ejemplo, una transformación que tiene efectos secundarios podría tener esta restricción.

Developer Tool habilita el método de optimización de predicado de forma predeterminada.

Método de optimización basado en el coste

Con una optimización basada en el coste, el Servicio de integración de datos evalúa una asignación, genera asignaciones semánticamente equivalentes y ejecuta la asignación con el rendimiento óptimo. La optimización basada en el coste reduce el tiempo de ejecución de asignaciones que realizan operaciones de combinación interior adyacentes y de combinación externa completa.

Las asignaciones semánticamente equivalentes son aquellas que realizan funciones idénticas y producen los mismos resultados. Para generar asignaciones semánticamente equivalentes, el Servicio de integración de datos divide en fragmentos la asignación original. Luego, el Servicio de integración de datos determina los fragmentos de asignación que puede optimizar.

Durante la optimización, el Servicio de integración de datos puede añadir, quitar o reordenar las transformaciones de un fragmento. El Servicio de integración de datos comprueba que los fragmentos optimizados produzcan los mismos resultados que los fragmentos originales y constituye asignaciones alternativas que utilizan los fragmentos optimizados.

El Servicio de integración de datos también puede aplicar una combinación de fusión ordenada si determina que el rendimiento de esta es mejor que el de una combinación de bucle anidado. Una combinación de fusión ordenada utiliza la ordenación para disponer dos conjuntos de datos antes de realizar la combinación. Una combinación de bucle anidado utiliza bucles anidados para combinar dos conjuntos de datos. El Servicio de integración de datos puede utilizar la información de ordenación en los orígenes o crear una transformación de ordenación si el coste de ordenar los datos es inferior al de procesar la combinación de bucle anidado.

El Servicio de integración de datos genera todas o prácticamente todas las asignaciones que son semánticamente equivalentes a la asignación original. Utiliza las estadísticas de base de datos o creación de perfiles para calcular el coste para la asignación original y para cada asignación alternativa. A continuación, identifica la asignación que se ejecuta más rápidamente. El Servicio de integración de datos realiza una

comprobación de validación en la mejor asignación alternativa para asegurar que es válida y que produce los mismos resultados que la asignación original.

El Servicio de integración de datos guarda en la memoria caché la mejor asignación alternativa. Cuando se ejecuta una asignación, el Servicio de integración de datos recupera la asignación alternativa y la ejecuta en lugar de la asignación original.

Developer tool no habilita este método de forma predeterminada.

Método de optimización dataship-join

El método de optimización dataship-join intenta buscar conjuntos de datos más pequeños junto a conjuntos de datos más grandes para reducir el tiempo de procesamiento de combinación. El Servicio de integración de datos intenta aplicar el método de optimización dataship-join cuando existe una importante diferencia de tamaño entre dos tablas.

Por ejemplo, el Servicio de integración de datos puede aplicar el método de optimización dataship-join para combinar una tabla principal con 10.000 filas con una tabla de detalles de 1.000.000 de filas. Para realizar la optimización dataship-join, el Servicio de integración de datos crea una tabla de ensayo temporal en la base de datos que contiene la tabla de detalles más grande. A continuación, el Servicio de integración de datos copia la tabla principal de menor tamaño en una tabla temporal y combina los datos de la tabla temporal con los datos de la tabla de detalles más grande. Después de que el Servicio de integración de datos realice la operación de combinación, la lógica de la transformación de combinación se procesa en la base de datos.

Antes de aplicar este método de optimización, el Servicio de integración de datos realiza varios análisis para determinar si la optimización dataship-join es factible y rentable. Si los análisis determinan que es probable que este método aumente el rendimiento, el Servicio de integración de datos lo aplica a la asignación. Luego, el Servicio de integración de datos vuelve a analizar la asignación para determinar si hay más oportunidades para realizar una optimización dataship-join. Si es necesario, lleva a cabo optimizaciones adicionales.

Developer tool no habilita este método de forma predeterminada.

Requisitos de dataship-join para mejorar el rendimiento

El método de optimización dataship-join no siempre aumenta el rendimiento. Los siguientes factores afectan al rendimiento de las asignaciones con la optimización dataship-join:

- El origen principal de la transformación de combinación debe tener muchas menos filas que el origen de detalle.
- El origen de detalle debe ser significativamente grande para justificar la optimización. Si el origen de detalle no es lo bastante grande, para el Servicio de integración de datos resulta más rápido leer todos los datos del origen principal y el de detalle sin aplicar el método de optimización dataship-join.

Reglas y directrices de optimización de dataship-join

El Servicio de integración de datos puede aplicar una optimización dataship-join a una transformación de combinación si la transformación cumple los siguientes requisitos:

- El tipo de combinación debe ser normal, outer principal u outer de detalle.
- El canal de detalle se debe originar en un origen relacional.
- Si la asignación utiliza commits basados en el destino, el ámbito de la transformación de combinación debe ser Todas las entradas.
- Los canales principales y de detalle no pueden compartir ninguna transformación.
- La asignación no puede contener una ramificación entre el origen de detalle y la transformación de incorporación.

- El Servicio de integración de datos fallará al aplicar el método de optimización dataship-join si la base de datos que contiene el lado de detalle de la combinación es una base de datos de IBM DB2 que no admite la codificación Unicode.

Método de optimización semi-join

El método de optimización semi-join reduce el número de datos extraídos del origen modificando operaciones de unión en la asignación.

El Servicio de integración de datos aplica el método de optimización semi-join a una transformación de combinación cuando, según la condición de combinación, un grupo de entrada tiene muchas más filas que el otro y cuando el grupo más grande tiene muchas más filas sin coincidencia en el grupo más pequeño. El Servicio de integración de datos intenta reducir el tamaño del conjunto de datos de un operando de unión leyendo las filas del grupo más pequeño, buscando las filas coincidentes en el grupo más grande y luego llevando a cabo la operación de unión. La reducción del tamaño del conjunto de datos mejora el rendimiento de la asignación, ya que el Servicio de integración de datos ya no debe leer filas innecesarias desde el origen del grupo más grande. El Servicio de integración de datos mueve la condición de combinación al origen del grupo más grande y solamente lee las filas que coinciden con el grupo más pequeño.

Antes de aplicar el método de optimización semi-join, el Servicio de integración de datos realiza varios análisis para determinar si la optimización semi-join es factible y rentable. Si los análisis determinan que es probable que este método aumente el rendimiento, el Servicio de integración de datos lo aplica a la asignación. Luego, el Servicio de integración de datos vuelve a analizar la asignación para determinar si hay más oportunidades para realizar una optimización semi-join. Si es necesario, lleva a cabo optimizaciones adicionales.

Developer Tool no habilita este método de forma predeterminada.

Requisitos de la optimización semi-join para aumentar el rendimiento

El método de optimización semi-join no siempre aumenta el rendimiento. Los siguientes factores afectan al rendimiento de las asignaciones con la optimización semi-join:

- El origen principal de la transformación de combinación debe tener muchas menos filas que el origen de detalle.
- El origen de detalle debe ser lo suficientemente grande para justificar la optimización. Cuando el servicio de integración de datos aplica la optimización semi-join, el método añade tiempo de sobrecarga al procesamiento de la asignación. Si el origen de detalle es pequeño, el tiempo necesario para aplicar el método semi-join puede sobrepasar el tiempo necesario para procesar todas las filas en el origen de detalle.
- El servicio de integración de datos debe poder obtener las estadísticas de recuento de filas para que una transformación de combinación pueda comparar con precisión las necesidades temporales de la operación de combinación normal con respecto a la operación semi-join.

Reglas y directrices de la optimización semi-join

El servicio de integración de datos puede aplicar una optimización semi-join a una transformación de combinación si la transformación cumple los siguientes requisitos:

- El tipo de unión debe ser normal, outer principal u outer de detalle. La transformación de incorporación no puede realizar un outer join completo.
- El canal de detalle se debe originar en un origen relacional.

- La condición de unión debe ser una condición ordenación-fusión-unión válida. Es decir, cada cláusula debe ser una igualdad de un puerto principal y un puerto de detalle. Si hay varias cláusulas, deben estar unidas por AND.
- Si la asignación no utiliza confirmaciones basadas en destino, el alcance de la transformación de incorporación debe ser Todas las entradas.
- Los canales principales y de detalle no pueden compartir ninguna transformación.
- La asignación no puede contener una ramificación entre el origen de detalle y la transformación de incorporación.

Método de optimización de primera selección

Cuando el Servicio de integración de datos aplica el método de optimización de primera selección, divide, mueve, o quita las transformaciones de filtro en una asignación. Mueve los filtros a un nivel superior de la asignación más cerca de origen.

El Servicio de integración de datos puede partir una transformación de filtro si la condición de filtro es una conjunción. Por ejemplo, el Servicio de integración de datos puede dividir la condición de filtro "A>100 AND B<50" en dos condiciones más simples: "A>100" y "B<50." Cuando el Servicio de integración de datos divide un filtro, mueve los filtros simplificados a un nivel superior del canal de la asignación, acercándolos al origen. El Servicio de integración de datos mueve los filtros a un nivel superior del canal por separado cuando se divide el filtro.

El método de optimización de primera selección está activado de forma predeterminada cuando elige el nivel normal o completo del optimizador en Developer tool. El Servicio de integración de datos ignora la optimización de primera selección si una transformación que aparece antes que la transformación de filtro tiene efectos secundarios. El Servicio de integración de datos no puede determinar si la transformación de SQL, la transformación del consumidor de servicio web y la transformación de Java tienen efectos secundarios. Puede configurar la optimización de primera selección para estas transformaciones si no tienen efectos secundarios.

Puede deshabilitar la primera selección si la optimización no aumenta el rendimiento. El Servicio de integración de datos habilita este método de optimización de forma predeterminada.

Método de optimización de predicado global

Cuando el Servicio de integración de datos utiliza el método de optimización de predicado global, elimina las filas que se pueden filtrar tan pronto como sea posible en la asignación. Esto reduce el número de filas que la asignación debe procesar. El método de optimización de predicado global incluye tanto la optimización de predicado como métodos de primera selección.

Por ejemplo, una asignación contiene una transformación de combinación con la condición de combinación "A=B" y una transformación de filtro con la condición de filtro "A>5". El Servicio de integración de datos puede agregar "B>5" a la condición de combinación y acercar la transformación de filtro al origen.

El método de optimización de predicado global aplica expresiones de predicado de manera más eficaz que el método de optimización de predicado. El método de optimización de predicado global determina si puede simplificar o volver a escribir las expresiones para aumentar el rendimiento de la asignación. También intenta aplicar expresiones de predicado tan pronto como sea posible en la asignación para mejorar el rendimiento de la asignación.

El método de optimización de predicado global deduce filtros y los acerca al origen cuando la asignación contiene ramificaciones o uniones anidadas con filtros en cada ramificación. Cuando el Servicio de integración de datos utiliza el método de optimización de predicado global, divide los filtros, los acerca al origen o los elimina en una asignación.

Método de optimización de eliminación de ramificaciones

El servicio de integración de datos puede aplicar el método de optimización de eliminación de ramificaciones a las transformaciones que no proporcionan ninguna fila al destino de una asignación.

El servicio de integración de datos puede quitar una transformación de filtro si la condición de filtro da como resultado FALSE para las filas de datos. Por ejemplo, una asignación tiene dos transformaciones de filtro que filtran datos de dos orígenes relacionales. Una transformación de filtro tiene la condición de filtro País=EE. UU. y la otra transformación de filtro tiene la condición de filtro País=Canadá. Una transformación de unión une los dos orígenes relacionales y tiene la condición de filtro País=EE. UU. El servicio de integración de datos puede quitar la transformación de filtro con la condición de filtro País=Canadá de la asignación.

La herramienta del desarrollador habilita el método de optimización de eliminación de ramificaciones de forma predeterminada cuando se selecciona el nivel normal o completo del optimizador. Puede deshabilitar la eliminación de ramificaciones si la optimización no aumenta el rendimiento cuando se establece el nivel del optimizador como mínimo o ninguno.

Método de la optimización de inserción

Con la optimización de inserción, el servicio de integración de datos mueve la lógica de transformación de filtro a la transformación del nivel inmediatamente superior de la transformación de filtro en la asignación. La optimización de inserción aumenta el rendimiento reduciendo el número de filas que pasan por la asignación.

El servicio de integración de datos no mueve la lógica de filtro a otra transformación si la transformación tiene efectos secundarios. El servicio de integración de datos no puede determinar si la transformación de SQL, la transformación del consumidor de servicio web y la transformación de Java tienen efectos secundarios. Puede configurar la transformación de SQL, la transformación del consumidor de servicio web y la transformación de Java para la optimización de inserción.

Optimización de inserciones

Cuando el servicio de integración de datos aplica la optimización de inserciones, inserta la lógica de transformación en la base de datos de origen. El servicio de integración de datos convierte la lógica de transformación en consultas SQL y las envía a la base de datos. La base de datos de origen ejecuta las consultas SQL para procesar las transformaciones.

La optimización de inserciones mejora el rendimiento de la asignación cuando la base de datos de origen puede procesar la lógica de transformación más rápidamente que el servicio de integración de datos. El servicio de integración de datos también lee menos datos del origen.

La cantidad de lógica de transformación que inserta el servicio de integración de datos en la base de datos de origen depende de la base de datos, la lógica de transformación y la configuración de la asignación. El servicio de integración de datos procesa toda la lógica de transformación que no puede insertar en una base de datos.

Cuando se aplica la optimización de inserciones, el servicio de integración de datos analiza la asignación optimizada desde el origen hasta el destino o hasta que alcance una transformación de nivel inferior que no pueda insertarse en la base de datos de origen. El servicio de integración de datos genera y ejecuta una consulta SELECT para cada origen que tiene una lógica de transformación insertada. El servicio de integración de datos también puede generar una consulta INSERT si el destino se ha insertado en la base de

datos. El servicio de integración de datos lee los resultados de las consultas SQL y procesa las transformaciones restantes de la asignación.

El servicio de integración de datos aplica la optimización de inserciones a una asignación al seleccionar el tipo de inserción en las propiedades de tiempo de ejecución de la asignación.

Puede seleccionar los siguientes tipos de inserción:

- Ninguno. No seleccione ningún tipo de inserción para la asignación.
- Origen. El servicio de integración de datos intenta insertar toda la lógica de transformación posible en la base de datos de origen.
- Completo. El servicio de integración de datos inserta la lógica de transformación completa en la base de datos de origen.

También puede crear un parámetro de cadena para el tipo de inserción y utilizar los siguientes valores de parámetro:

- Ninguno
- Origen
- Completo

Optimización de inserciones completa

Cuando el servicio de integración de datos aplica la optimización de inserciones completa, inserta toda la lógica de transformación de la asignación en la base de datos de origen. Puede configurar la inserción completa en las propiedades de tiempo de ejecución de la asignación.

La optimización de inserciones completa es ideal cuando el origen y el destino se encuentran en la misma base de datos o cuando las transformaciones, como las de agregación y de filtro, se procesan en la base de datos de origen y reducen la cantidad de datos movidos. Por ejemplo, si una asignación contiene un origen y un destino de Teradata, configure la optimización de inserciones completa para insertar toda la lógica de la transformación para el procesamiento de una base de datos de origen de Teradata a una base de datos de destino de Teradata.

Al configurar una asignación con una transformación de estrategia de actualización para la inserción completa, debe determinar la compatibilidad de inserción de la asignación.

El servicio de integración de datos puede insertar una asignación con una transformación de estrategia de actualización en los siguientes casos:

- La transformación de destino conectada a la transformación de estrategia de actualización recibe varias filas que no tienen la misma clave.
- La transformación de destino conectada a la transformación de estrategia de actualización recibe varias filas con la misma clave que se pueden reordenar.

El servicio de integración de datos no puede insertar una asignación con una transformación de estrategia de actualización en el siguiente caso:

- La transformación de destino conectada a la transformación de estrategia de actualización recibe varias filas con la misma clave que no se pueden reordenar.

También puede utilizar un parámetro de compatibilidad de inserción en la asignación. Puede usar los siguientes valores de parámetro:

- noMultipleRowsWithSameKeyOnTarget
- reorderAllowedForMultipleRowsWithSameKey
- reorderNotAllowedForRowsWithSameKey

El servicio de integración de datos puede utilizar la optimización de inserciones completa con los siguientes orígenes:

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

Inserción de origen

Cuando el Servicio de integración de datos aplica la inserción de origen, analiza la asignación del origen al destino o hasta que llega a una transformación de nivel inferior que no puede insertar en la base de datos de origen.

El Servicio de integración de datos genera y ejecuta una declaración SELECT basándose en la lógica de transformación para cada transformación que puede insertar en la base de datos. A continuación, lee los resultados de esta consulta SQL y procesa las transformaciones restantes.

Puede configurar una asignación para utilizar la inserción de origen si el origen y el destino residen en distintas bases de datos. Por ejemplo, si una asignación contiene un origen de Teradata y un destino de Oracle, puede configurar la inserción de origen para insertar lógica de transformación para el procesamiento en el origen de Teradata.

Reglas y directrices de la optimización de inserciones

El servicio de integración de datos puede insertar lógica de transformación en la base de datos de origen.

Las siguientes reglas y directrices se aplican a la optimización de inserciones:

- El servicio de integración de datos puede insertar la lógica de transformación de búsqueda y combinación a la base de datos de origen si los orígenes están en el mismo sistema de administración de bases de datos y utilizan conexiones idénticas.
- El servicio de integración de datos no puede insertar la lógica de transformación en un origen que tiene un tipo de datos binario.
- El servicio de integración de datos inhabilita la optimización de inserciones cuando existe un origen de datos IBM DB2 y la precisión de columna para el tipo de datos decimal está entre 28 y 31 dígitos.
- El servicio de integración de datos habilita la optimización de inserciones para un servicio de datos SQL o un servicio web de forma predeterminada. No se puede inhabilitar la optimización de inserciones para un servicio de datos SQL o un servicio web.
- El servicio de integración de datos no puede insertar una transformación de agregación que contenga una expresión con funciones agregadas y no agregadas en un puerto que no sea de agrupación.

Lectura de pase individual

La lectura de pase individual le permite llenar múltiples destinos con un objeto de datos personalizado. Considere utilizar la lectura de pase individual si tiene varias asignaciones que utilizan los mismos orígenes.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de la lectura de pase individual:

Combine la lógica de transformación para cada asignación en una asignación y utilice un objeto de datos personalizado para cada origen.

El servicio de integración de datos lee cada origen una vez y, a continuación, envía los datos en distintos canales. Una fila determinada puede ser utilizada por todos los canales, por cualquier combinación de canales o por ningún canal.

Por ejemplo, tiene la tabla de origen de compras y utiliza ese origen a diario para realizar una agregación y una clasificación. Si coloca las transformaciones de agregación y de rango en asignaciones separadas, puede forzar al servicio de integración de datos a leer la misma tabla de origen dos veces. Sin embargo, si incluye la lógica de agregación y de clasificación en una asignación con un calificador de origen, el servicio de integración de datos lee la tabla de origen de compras una vez y, a continuación, envía los datos correspondientes a los distintos canales.

Excluya funciones comunes de las asignaciones.

Al cambiar asignaciones para aprovechar las prestaciones de la lectura de pase individual, puede optimizar esta característica por excluyendo las funciones comunes de las asignaciones. Por ejemplo, si necesita restar un porcentaje de los puertos de Precio para las transformaciones tanto de agregación como de rango, puede reducir el trabajo restando el porcentaje antes de dividir el canal. Puede utilizar una transformación de expresión para restar el porcentaje y, a continuación, dividir la asignación después de la transformación.

Optimización del filtro

Puede optimizar las asignaciones filtrando dentro del objeto de datos personalizado y colocando filtros al principio de la asignación.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de filtros:

Utilice un filtro en un objeto de datos personalizado para quitar las filas en el origen.

Si filtra filas desde la asignación, puede mejorar la eficacia filtrando primero el flujo de datos. Utilice un filtro en un objeto de datos personalizado para quitar las filas en el origen. El objeto de datos personalizado limita el conjunto de filas extraído de un origen relacional.

Si no puede utilizar un filtro en el objeto de datos personalizado, utilice una transformación de filtro y muévela tan cerca del objeto de datos personalizado como sea posible para quitar los datos innecesarios al principio del flujo de datos. La transformación de filtro limita el conjunto de filas enviadas a un destino.

Utilice un filtro en una transformación de estrategia de actualización si no necesita mantener las filas rechazadas.

Para mejorar el rendimiento de la asignación, también puede utilizar una transformación de filtro para quitar las filas rechazadas desde una transformación de estrategia de actualización si no necesita mantener las filas rechazadas.

Evitar expresiones complejas en las condiciones de filtro.

Evitar el uso de expresiones complejas en las condiciones de filtro. Para optimizar las transformaciones de filtro, utilice un entero simple o expresiones TRUE/FALSE en la condición del filtro.

La transformación de filtro filtra datos dentro de una asignación. La transformación de filtro filtra filas desde cualquier tipo de origen. El objeto de datos personalizado filtra las filas de los orígenes relacionales. La transformación de filtro filtra filas desde cualquier tipo de origen.

Optimización de conversión de tipos de datos

Puede aumentar el rendimiento eliminando conversiones de tipos de datos innecesarias. Por ejemplo, si una asignación mueve datos desde una columna de enteros a una columna de decimales y, a continuación, de nuevo a la columna de enteros, la conversión de tipos de datos innecesaria disminuye el rendimiento. Si es posible, elimine conversiones de tipo de datos innecesarias desde las asignaciones.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de las conversiones de tipos de datos: **Utilice valores enteros en lugar de otros tipos de datos cuando realice comparaciones mediante la búsqueda y las transformaciones de filtro.**

Por ejemplo, muchas las bases de datos almacenan información de código postal de Estados Unidos como un tipo de datos Char o Varchar. Si convierte los datos del código postal a un tipo de datos entero, la base de datos de búsqueda almacena el código postal 94303-1234 como 943031234. Esto ayuda a aumentar la velocidad de las comparaciones de búsqueda basándose en el código postal.

Convierta las fechas de origen a cadenas a través de las conversiones puerto a puerto para aumentar el rendimiento de la asignación.

Puede dejar los puertos en los destinos como cadenas o cambiarlos a puertos a Fecha/Hora.

Error de seguimiento

Para mejorar el rendimiento, reduzca el número de eventos de registro generados por el servicio de integración de datos cuando ejecuta la asignación. Mejore el rendimiento de las asignaciones actualizando el nivel del optimizador de asignaciones mediante la configuración de la asignación o las propiedades de implementación de la asignación. Utilice el método de optimización basada en el coste para optimizar las asignaciones.

Tenga en cuenta las siguientes soluciones para los cuellos de botella del error de seguimiento:

Establezca el nivel de seguimiento en las propiedades de la asignación en simplificado

Si una asignación contiene un gran número de errores de transformación y no necesita corregirlos, establezca el nivel de seguimiento en las propiedades de la asignación en simplificado. En este nivel de seguimiento, el servicio de integración de datos no escribe mensajes de error o información a nivel de fila para rechazar datos.

Si necesita depurar la asignación y establece el nivel de seguimiento en detallado, es posible que experimente una degradación considerable del rendimiento cuando ejecute la asignación. No utilice el seguimiento detallado al ajustar el rendimiento. El nivel de seguimiento de la asignación reemplaza cualquier nivel de seguimiento específico de la transformación dentro de la asignación. Esto no se recomienda como una respuesta a largo plazo para altos niveles de errores de transformación.

Cambie el nivel del optimizador para la asignación.

Si una asignación tarda un tiempo excesivo en ejecutarse, sería recomendable cambiar el nivel del optimizador para la asignación. El nivel del optimizador determina los métodos de optimización que el servicio de integración de datos aplica a la asignación en tiempo de ejecución.

El nivel del optimizador para una asignación se define en la configuración de la asignación o en las propiedades de implementación de la asignación. El servicio de integración de datos aplica distintos niveles del optimizador a la asignación según cómo se ejecute la asignación.

Utilice el método de optimización basado en el coste.

El método de optimización basado en el coste hace que el servicio de integración de datos evalúe una asignación, genere asignaciones semánticamente equivalentes y ejecute la asignación con el mayor rendimiento posible. Este método es el más eficaz para las asignaciones que contienen varias transformaciones de incorporación. Reduce el tiempo de ejecución de asignaciones que llevan a cabo operaciones de combinación interior adyacentes y no ordenadas.

Las asignaciones semánticamente equivalentes son aquellas que realizan funciones idénticas y producen los mismos resultados. Para generar asignaciones semánticamente equivalentes, el servicio de integración de datos divide en fragmentos la asignación original. Luego, el servicio de integración de datos determina los fragmentos de asignación que puede optimizar.

CAPÍTULO 6

Optimización de asignación con particiones

Este capítulo incluye los siguientes temas:

- [Resumen de optimización de asignación con particiones, 55](#)
- [Utilizar varias CPU, 56](#)
- [Aumentar el número máximo de paralelismos, 56](#)
- [Optimizar archivos sin formato para particiones, 57](#)
- [Optimizar las bases de datos relacionales para la partición, 58](#)
- [Optimizar transformaciones para particiones, 59](#)

Resumen de optimización de asignación con particiones

Si tiene la opción de partición, puede habilitar el servicio de integración de datos para maximizar el paralelismo cuando ejecute asignaciones. Cuando se maximiza el paralelismo, el servicio de integración de datos divide de forma dinámica los datos subyacentes en particiones y procesa todas las particiones simultáneamente.

Si las asignaciones procesan grandes conjuntos de datos o contienen transformaciones que realizan cálculos complicados, las asignaciones pueden tardar mucho tiempo en procesarse y pueden provocar un bajo rendimiento de los datos. Cuando se habilita la partición para estas asignaciones, el servicio de integración de datos utiliza subprocesos adicionales para procesar la asignación.

Puede optimizar el rendimiento de asignaciones con particiones mediante las tareas siguientes:

- Utilice varias CPU en los nodos que ejecutan asignaciones.
- Aumente el número máximo de paralelismos para el servicio de integración de datos.
- Configure propiedades de los objetos de datos de archivo sin formato.
- Configure bases de datos relacionales para optimizar las particiones.
- Configure propiedades en transformaciones.

Utilizar varias CPU

Aumentar el número de subprocesos de procesamiento aumenta la carga en los nodos que ejecutan asignaciones. Si los nodos contienen un ancho de banda de CPU amplio, el procesamiento simultáneo de filas de datos en una asignación puede optimizar el rendimiento de la asignación.

El Servicio de integración de datos puede utilizar varias CPU para procesar una asignación que contiene varias particiones. El número de CPU que el servicio utiliza depende de factores como el número de puntos de partición, el número de subprocesos creados para cada etapa de canal y la cantidad de recursos necesarios para procesar la asignación. Una asignación simple se ejecuta más rápidamente en dos particiones, pero suele requerir el doble de CPU que cuando se ejecuta la asignación en una única partición.

Aumentar el número máximo de paralelismos

El número máximo de paralelismos determina la cantidad máxima de subprocesos paralelos que puede procesar una única etapa de canal. Configure la propiedad **Número máximo de paralelismos** para el Servicio de integración de datos basándose en los recursos de hardware disponibles. Cuando aumente el valor del número máximo de paralelismos, podrá reducir la cantidad de tiempo de procesamiento.

Tenga en cuenta las siguientes directrices cuando aumente el valor del número máximo de paralelismos:

Aumente el valor según el número de CPU disponibles.

Aumente el valor del número máximo de paralelismos en función del número de CPU disponibles en los nodos donde se ejecutan las asignaciones. Al aumentar el valor del número máximo de paralelismos, el Servicio de integración de datos utiliza más subprocesos para ejecutar la asignación y aprovecha más CPU. Una asignación simple se ejecuta más rápidamente en dos particiones, pero suele requerir el doble de CPU que cuando se ejecuta la asignación en una única partición.

Tenga en cuenta el número total de subprocesos de procesamiento.

Tenga en cuenta el número total de subprocesos de procesamiento cuando establezca el valor del número máximo de paralelismos. Si una asignación compleja produce varios puntos de partición adicionales, el Servicio de integración de datos podría utilizar más subprocesos de procesamiento de los que puede manejar la CPU.

El número total de subprocesos de procesamiento es igual al valor del número máximo de paralelismos.

Tenga en cuenta los otros trabajos que debe ejecutar el Servicio de integración de datos.

Si configura el número máximo de paralelismos de tal forma que cada asignación utilice un gran número de subprocesos, quedarán menos subprocesos disponibles para que el Servicio de integración de datos ejecute tareas adicionales.

También puede cambiar el valor para una asignación.

De manera predeterminada, el número máximo de paralelismos de cada asignación está establecido en Automático. Todas las asignaciones utilizan el valor del número máximo de paralelismos definido para el Servicio de integración de datos.

En Developer tool, los desarrolladores pueden cambiar el valor del número máximo de paralelismos en las propiedades de tiempo de ejecución de la asignación para definir un valor máximo para una asignación concreta. Cuando el número máximo de paralelismos está establecido en valores enteros diferentes para el Servicio de integración de datos y para la asignación, el Servicio de integración de datos utiliza el valor más pequeño de los dos.

Optimizar archivos sin formato para particiones

Cuando una asignación habilitada para particiones lee desde un origen del archivo sin formato o escribe en un destino del archivo sin formato, el servicio de integración de datos puede utilizar varios subprocesos para leer o para escribir en el archivo sin formato.

Optimizar orígenes de archivos sin formato para particiones

A fin de obtener un rendimiento óptimo al utilizar varios subprocesos para leer de un archivo sin formato, configure el objeto de datos de archivo sin formato para optimizar el rendimiento en lugar de conservar el orden de las filas.

Considere la siguiente solución para reducir los cuellos de botella de los orígenes de archivos sin formato con particiones:

Configure la partición de lectura simultánea para el objeto de datos de archivo sin formato para optimizar el rendimiento.

En las propiedades avanzadas del objeto de datos de archivo sin formato, establezca la propiedad **Partición de lectura simultánea** para optimizar el rendimiento. Al optimizar el rendimiento, el Servicio de integración de datos no conserva el orden de las filas porque no lee las filas del archivo o la lista de archivos de forma secuencial.

Optimizar destinos de archivos sin formato para particiones

A fin de obtener un rendimiento óptimo al utilizar varios subprocesos para escribir en un archivo sin formato, configure particiones para escribir la salida de destino en archivos distintos y configurar varios directorios de destino.

Considere las siguientes soluciones para reducir los cuellos de botella de los destinos de archivos sin formato con particiones:

Configure particiones para escribir la salida de destino en archivos distintos.

En las propiedades avanzadas del objeto de datos de archivo sin formato, establezca la propiedad **Tipo de fusión en Sin fusión**. El Servicio de integración de datos escribe simultáneamente la salida de destino en un archivo distinto para cada partición. Si necesita datos de destino fusionados, el tipo de fusión simultánea optimizar el rendimiento más que el tipo de fusión secuencial.

Configure varios directorios de destino.

Cuando varios subprocesos escriben en un solo directorio, la asignación podría experimentar un cuello de botella debido a la contención de E/S. Una contención de E/S puede producirse cuando varios subprocesos escriben datos en el sistema de archivos al mismo tiempo. Al configurar varios directorios, el Servicio de integración de datos determina el directorio de salida para cada subproceso mediante un método Round-Robin.

Configure los directorios de archivo de salida en las propiedades avanzadas del objeto de datos de archivo sin formato. Utilice el valor del parámetro del sistema predeterminado de TargetDir si un administrador ha introducido varios directorios separados por punto y coma para la propiedad **Directorio de destino** para el Servicio de integración de datos en la Herramienta del administrador. También puede introducir un valor distinto para configurar varios directorios de archivo de salida específicos para el objeto de datos de archivo sin formato.

Optimizar las bases de datos relacionales para la partición

Cuando una asignación habilitada para particiones lee o escribe en una base de datos relacional DB2 para LUW de IBM o en una base de datos relacional de Oracle, el servicio de integración de datos puede utilizar varios subprocesos para leer el origen relacional o escribir en el destino relacional.

Para lograr un rendimiento óptimo cuando se utilizan varios subprocesos para leer desde una base de datos relacional de DB2 para LUW o de Oracle o escribir en ella, puede realizar una partición de las tablas de origen y de destino.

Nota: Si una asignación lee o escribe en una base de datos relacional que no sea DB2 para LUW u Oracle, el servicio de integración de datos utiliza un subproceso de lector o de escritor.

Optimizar la base de datos de origen para la partición

Para lograr un rendimiento óptimo cuando se utilizan varios subprocesos para leer desde una base de datos de origen de DB2 para LUW o de Oracle, compruebe que la tabla de origen tiene particiones y está configurada para aceptar consultas en paralelo.

Para optimizar la base de datos de origen para la partición, realice las tareas siguientes:

Añadir particiones de la base de datos al origen.

Añada particiones de la base de datos al origen relacional para aumentar la velocidad de la consulta del servicio de integración de datos que lee el origen. Si el origen no tiene particiones de la base de datos, el servicio de integración de datos utiliza un subproceso para leer en el origen.

Habilitar consultas en paralelo.

Las bases de datos relacionales pueden tener opciones que habiliten las consultas en paralelo en la base de datos. Consulte la información sobre estas opciones en la documentación de la base de datos. Si estas opciones no están habilitadas, el servicio de integración de datos ejecuta varias instrucciones SELECT de particiones en serie.

Separar datos en diferentes espacios de tablas.

Cada base de datos proporciona una opción para separar los datos en diferentes espacios de tablas. Cada espacio de tablas puede hacer referencia a un sistema de archivo único, lo que impide que haya contenciones de E/S en las particiones.

Aumentar el número máximo de sesiones permitidas en la base de datos.

El servicio de integración de datos crea una conexión con la base de datos de origen distinta para cada partición. Aumente el número máximo de sesiones permitidas para que la base de datos pueda manejar un número mayor de conexiones simultáneas.

Optimizar la base de datos de destino para la partición

Para lograr un rendimiento óptimo cuando se utilizan varios subprocesos para escribir en una base de datos de destino de DB2 para LUW o de Oracle, asegúrese de que la tabla de destino tiene particiones y está configurada para insertar filas en paralelo.

Para optimizar la base de datos de destino para la partición, realice las tareas siguientes:

Añadir particiones de la base de datos a un destino de DB2 para LUW.

El servicio de integración de datos puede utilizar varios subprocesos para escribir en un destino de DB2 para LUW que no tiene particiones de la base de datos. No obstante, puede optimizar el rendimiento de

carga cuando el destino tiene particiones de la base de datos. En este caso, cada subproceso de escritura se conecta al nodo de DB2 para LUW que contiene la partición de la base de datos. El rendimiento aumenta porque los subprocesos de escritura se conectan a distintos nodos de DB2 para LUW, en lugar de que se conecten todos los subprocesos al único nodo principal.

Habilitar inserciones en paralelo.

Las bases de datos relacionales pueden tener opciones que habiliten las inserciones en paralelo en la base de datos. Consulte la información sobre estas opciones en la documentación de la base de datos. Por ejemplo, establezca la opción `db_writer_processes` en una base de datos Oracle y la opción `max_agents` en una base de datos de DB2 para LUW para habilitar las inserciones en paralelo.

Separar datos en diferentes espacios de tablas.

Cada base de datos proporciona una opción para separar los datos en diferentes espacios de tablas. Cada espacio de tablas puede hacer referencia a un sistema de archivo único, lo que impide que haya contenciones de E/S en las particiones.

Aumentar el número máximo de sesiones permitidas en la base de datos.

El servicio de integración de datos crea una conexión con la base de datos de destino distinta para cada partición. Aumente el número máximo de sesiones permitidas para que la base de datos pueda manejar un número mayor de conexiones simultáneas.

Establecer opciones para mejorar la escalabilidad de la base de datos.

Las bases de datos relacionales pueden tener opciones que mejoran la escalabilidad. Por ejemplo, deshabilite el registro de los archivos y la estadísticas temporizadas en una base de datos de Oracle para mejorar la escalabilidad.

Optimizar transformaciones para particiones

Cuando el Servicio de integración de datos utiliza varios subprocesos para ejecutar una transformación de agregación, combinación, rango u ordenación, el servicio usa la partición de la memoria caché para dividir el tamaño de la memoria caché entre los subprocesos. Para optimizar el rendimiento para la partición de la memoria caché, configure varios directorios de la memoria caché.

Nota: Una transformación de búsqueda solo puede utilizar un único directorio de la memoria caché.

Considere la siguiente solución para reducir los cuellos de botella de las transformaciones de agregación, combinación, rango u ordenación:

Configure varios directorios de la memoria caché.

La partición de la memoria caché crea una memoria caché diferente para cada partición que procese una transformación de agregación, combinación, rango u ordenación. Durante la partición de la memoria caché, cada partición almacena datos diferentes en una memoria caché distinta. Cada memoria caché contiene las filas que necesita esa partición. La partición de la memoria caché optimiza el rendimiento de la asignación, porque cada subproceso consulta una memoria caché diferente en paralelo.

Si el tamaño de la memoria caché es menor que la cantidad de memoria necesaria para ejecutar la transformación, los subprocesos de transformación escriben en el directorio de la memoria caché para almacenar valores de desbordamiento en archivos de memoria caché. Cuando varios subprocesos escriben en un mismo directorio, la asignación podría experimentar un cuello de botella debido a la contención de E/S. Una contención de E/S puede producirse cuando varios subprocesos escriben datos en el sistema de archivos al mismo tiempo. Cuando configura varios directorios de memoria caché, el

Servicio de integración de datos determina el directorio de memoria caché para cada subproceso de transformación de modo Round-Robin.

En una transformación de agregación, combinación o rango, configure los directorios de memoria caché en la propiedad avanzada **Directorio de la memoria caché**. Utilice el valor del parámetro del sistema predeterminado de CacheDir si un administrador ha introducido varios directorios separados por punto y coma para la propiedad **Directorio de la memoria caché** para el Servicio de integración de datos en la Herramienta del administrador. También puede introducir un valor distinto para configurar varios directorios de memoria caché específicos para la transformación.

En una transformación de ordenación, configure los directorios de memoria caché en la propiedad avanzada **Directorio de trabajo**. Utilice el valor del parámetro del sistema predeterminado de TempDir si un administrador ha introducido varios directorios separados por punto y coma para la propiedad **Directorios temporales** para el Servicio de integración de datos en la Herramienta del administrador. También puede introducir un valor distinto para configurar varios directorios de memoria caché específicos para la transformación.

CAPÍTULO 7

Optimización del tiempo de ejecución

Este capítulo incluye los siguientes temas:

- [Resumen de optimización del tiempo de ejecución, 61](#)
- [Optimización del servicio de aplicación, 61](#)
- [Estadísticas de supervisión, 64](#)
- [Asignación de memoria, 65](#)
- [Memoria caché de objetos de datos, 66](#)
- [Optimización del sistema, 69](#)

Resumen de optimización del tiempo de ejecución

Habilite las características de rendimiento y ajuste las propiedades de Servicio de integración de datos para optimizar el rendimiento de las asignaciones.

Utilice las siguientes técnicas de optimización en Administrator Tool para obtener los mejores resultados de rendimiento según sus requisitos:

- Optimice los procesos de servicio de aplicación.
- Configure las estadísticas de supervisión para supervisar los cuellos de botella del sistema.
- Asigne memoria para obtener un rendimiento óptimo del sistema.
- Configure la memoria caché de objetos de datos.
- Optimice el sistema para evitar retrasos del sistema y reducir el acceso al disco.

Optimización del servicio de aplicación

Optimice el proceso de servicio de la aplicación cuando el rendimiento se vea afectado. Puede optimizar el servicio del analista, el servicio de integración de datos y el servicio de repositorio de modelos.

Optimización del servicio del analista

Ajuste el servicio del analista para optimizar el rendimiento. Puede configurar las propiedades del proceso del servicio del analista para la memoria, minimizar la latencia de la red y establecer la configuración de carga del archivo sin formato de Analyst Tool para mejorar el rendimiento del servicio.

Tenga en cuenta las siguientes soluciones para los cuellos de botella del servicio del analista:

Configure Analyst Tool para que se conecte a una ubicación de ruta de acceso a la red para cargar archivos sin formato de más de 10 MB.

El rendimiento de proceso de servicio del analista se puede reducir cuando los analistas cargan archivos sin formato grandes de más de 10 MB en el directorio de instalación de Informatica en el equipo en el que se ejecuta la herramienta del analista. Esto puede afectar tanto al espacio en disco como al rendimiento de la red.

Cargue los archivos sin formato de menos de 10 MB en el directorio de instalación de Informatica desde Analyst Tool.

El rendimiento del proceso del servicio del analista se puede reducir si los analistas cargan archivos sin formato de más de 10 MB en el directorio de instalación de Informatica desde Analyst Tool. Esto puede afectar tanto al espacio en disco como al rendimiento de la red.

Aumente la propiedad del tamaño de heap máximo para el proceso del servicio del analista.

El proceso del servicio de analista puede consumir grandes cantidades de memoria durante el procesamiento de un gran número de usuarios conectados de forma simultánea. Esto puede provocar que se abra un gran número de conexiones de red entre el servicio del analista y otros servicios como el servicio de integración de datos o el servicio de repositorio de modelos.

Utilice Administrator Tool para configurar la propiedad del tamaño de heap máximo en un valor mayor en las Propiedades avanzadas del proceso del servicio del analista.

Exporte las especificaciones de asignación grandes a una tabla o expórtelas a un archivo sin formato y trunque el archivo.

El proceso de servicio del analista puede tener un impacto en el rendimiento cuando los analistas exportan grandes especificaciones de asignación como archivos sin formato desde Analyst Tool.

Optimización del servicio de integración de datos

Ajuste el proceso del servicio de integración de datos para mejorar el rendimiento del servicio. Puede configurar las propiedades del proceso del servicio de integración de datos para la memoria. Puede configurar cada servicio web y servicio de datos SQL que se ejecuta en un servicio de integración de datos para gestionar solicitudes simultáneas.

Tenga en cuenta las siguientes soluciones para los cuellos de botella del servicio de integración de datos:

Configure la propiedad tamaño de heap máximo para el proceso del servicio de integración de datos.

El servicio de integración de datos puede consumir grandes cantidades de memoria durante el procesamiento de los servicios de datos SQL y los servicios web.

Use Administrator Tool para configurar la propiedad del tamaño de heap máximo en un valor mayor en las propiedades avanzadas para el proceso del servicio de integración de datos.

Configure la propiedad Tiempo para mantener DTM del servicio del web para el servicio de integración de datos.

El servicio de integración de datos consume los recursos del sistema para generar una instancia del DTM para cada solicitud de servicio web. Configure el servicio de integración de datos para usar una instancia del DTM para procesar más de una solicitud de servicio web.

Use Administrator Tool para configurar la propiedad Tiempo para mantener DTM del servicio web para el servicio de integración de datos.

Configure las opciones de ejecución en las propiedades del proceso de integración de datos y el servicio web y las propiedades del servicio de datos SQL para solicitudes simultáneas.

El servicio de integración de datos, cada servicio de datos SQL y cada servicio web que se ejecuta en el servicio de integración de datos consume recursos del sistema y de la memoria para cada solicitud simultánea.

Para configurar el número de solicitudes simultáneas que pueden aceptar el servicio de integración de datos, cada servicio de datos SQL y cada servicio web, configure las propiedades del servicio de integración de datos y las propiedades del servicio web.

Use la herramienta Administrator para configurar las siguientes opciones y propiedades del servicio de integración de datos, el servicio web y el servicio de datos SQL:

- Configure las opciones de ejecución del servicio de integración de datos.
- Configure la propiedad Número máximo de conexiones simultáneas para cada servicio de datos SQL en las propiedades SQL para el proceso del servicio de integración de datos.
- Configure las propiedades del número máximo de solicitudes acumuladas y del número máximo de solicitudes simultáneas para cada servicio web en las propiedades de configuración de HTTP para el proceso del servicio de integración de datos.

Desactive el nivel de seguimiento del servicio web.

El número de archivos de registro del servicio web que el servicio de integración de datos escribe y mantiene puede reducir el rendimiento.

Use la herramienta Administrator para configurar el nivel de seguimiento del servicio web para reducir la cantidad de archivos de registro en tiempo de ejecución del servicio web que el servicio de integración de datos almacena en el disco.

Optimización del servicio de repositorio de modelos

Ajuste el servicio de repositorio de modelos para mejorar el rendimiento. Puede configurar la propiedad del proceso del servicio de repositorio de modelos para la memoria y minimizar la latencia de red.

Tenga en cuenta las siguientes soluciones a los cuellos de botella del servicio de repositorio de modelos:

Aloje la base de datos del repositorio de modelos en el mismo equipo que el servicio de repositorio de modelos.

El rendimiento del proceso del servicio de repositorio de modelos puede verse afectado si la base de datos del repositorio de modelos está alojada en un servidor remoto. Las operaciones del servicio de repositorio de modelos que requieren comunicación entre el repositorio de modelos y el servicio de repositorio de modelos en una red de alta latencia podría ralentizar el rendimiento del servicio de repositorio de modelos.

Aumente la propiedad tamaño de heap máximo para el proceso del servicio de repositorio de modelos.

El proceso del servicio de repositorio de modelos puede consumir grandes cantidades de memoria durante el procesamiento de un gran número de usuarios registrados simultáneamente. Esto puede provocar que haya que abrir un gran número de conexiones de red entre el servicio de repositorio de modelos y otros servicios como el servicio de integración de datos o el servicio del analista.

Utilice la Herramienta del administrador para configurar la propiedad del tamaño de heap máximo en un valor mayor en las propiedades avanzadas para el proceso del servicio de repositorio de modelos.

Estadísticas de supervisión

La supervisión es una función del dominio realizada por el administrador de servicios. El administrador de servicios almacena la configuración de supervisión en el repositorio de modelos. Utilice la ficha Supervisión en la Herramienta del administrador para supervisar los cuellos de botella del sistema, tales como el número total de tareas en ejecución, que han fallado, canceladas y finalizadas que se ejecutan en un servicio seleccionado.

Tenga en cuenta la siguiente solución para supervisar los cuellos de botella de estadísticas:

Configure el dominio para configurar la supervisión.

Al configurar la supervisión, el Servicio de integración de datos almacenará las estadísticas persistidas y los informes de supervisión en el repositorio de modelos. Las estadísticas persistidas son un historial de información sobre objetos de integración que se han ejecutado previamente. Los informes de supervisión muestran mediciones clave sobre un objeto de integración.

Configure valores de supervisión para el dominio a fin de especificar el repositorio de modelos que almacena las estadísticas de tiempo de ejecución sobre los objetos implementados en los servicios de integración de datos. Los valores de supervisión se aplican a todos los servicios de integración de datos del dominio y pueden afectar al rendimiento del servicio.

La tabla siguiente describe los valores de supervisión que pueden afectar al rendimiento de los servicios:

Opción	Descripción
Conservar datos históricos de resumen	Número de días que el repositorio de modelos guarda datos de promedio. Si la purga está deshabilitada, el repositorio de modelos guarda los datos indefinidamente. El valor predeterminado es 180. El valor mínimo es 0. El valor máximo es 366.
Conservar datos históricos detallados	Número de días que el repositorio de modelos guarda datos por minuto. Si la purga está deshabilitada, el repositorio de modelos guarda los datos indefinidamente. El valor predeterminado es 14. El valor mínimo es 1. El valor máximo es 14.
Purgar estadísticas cada	Intervalo, en días, en el que el servicio de repositorio de modelos purga datos anteriores a los valores configurados en las opciones de conservación de datos históricos . El valor predeterminado es 1 día.
Días en	Hora del día en que el servicio de repositorio de modelos purgará las estadísticas. El valor predeterminado es 1:00 a. m.
Número máximo de registros ordenables	Número máximo de registros que se pueden ordenar en la ficha Supervisar . Si el número de registros de la ficha Supervisar es mayor que este valor, solo podrá ordenar por Hora de inicio y Hora de finalización . El valor predeterminado es 3.000.
Retardo máximo para las notificaciones de actualización	Tiempo máximo, en segundos, durante el cual el servicio de integración de datos almacena en la memoria intermedia las estadísticas antes de almacenarlas en el repositorio de modelos y de mostrarlas en la ficha Supervisar . Si el servicio de integración de datos se cierra inesperadamente antes de que almacene las estadísticas en el repositorio de modelos, las estadísticas se perderán. El valor predeterminado es 10.
Mostrar milisegundos	Incluye los milisegundos para los campos de fecha y hora en la ficha Supervisión .

Asignación de memoria

Para optimizar el rendimiento de las asignaciones, configure las propiedades de memoria para el Servicio de integración de datos en la Herramienta del administrador.

En la siguiente tabla se describe la propiedad Memoria máxima por solicitud para el Módulo del servicio de asignación:

Propiedad	Descripción
Memoria máxima por solicitud	<p>El comportamiento de Memoria máxima por solicitud depende de las siguientes configuraciones del servicio de integración de datos:</p> <ul style="list-style-type: none">- El servicio ejecuta trabajos en procesos locales o remotos independientes, o la propiedad del servicio Tamaño máximo de memoria es 0 (predeterminado). <p>En este caso, Memoria máxima por solicitud es la cantidad máxima de memoria, en bytes, que el servicio de integración de datos puede asignar a todas las transformaciones que utilizan el modo de memoria caché automática en una sola solicitud. El servicio asigna la memoria por separado a las transformaciones que tienen un tamaño de memoria caché específico. La memoria total utilizada por la solicitud puede superar el valor de Memoria máxima por solicitud.</p> <ul style="list-style-type: none">- El servicio ejecuta trabajos en el proceso de servicio de integración de datos, y la propiedad del servicio Tamaño máximo de memoria es mayor que 0. <p>En este caso, Memoria máxima por solicitud es la cantidad máxima de memoria, en bytes, que el servicio de integración de datos puede asignar a una sola solicitud. La memoria total utilizada por la solicitud no puede superar el valor de Memoria máxima por solicitud.</p> <p>El valor predeterminado es 536 870 912.</p>

En la siguiente tabla se describen las opciones de ejecución para el Servicio de integración de datos:

Propiedad	Descripción
Tamaño máximo de memoria	<p>Cantidad máxima de memoria, en bytes, que el servicio de integración de datos puede asignar para ejecutar todas las solicitudes de forma simultánea cuando el servicio ejecuta trabajos en el proceso de servicio de integración de datos. Cuando el servicio de integración de datos ejecuta trabajos en procesos locales o remotos independientes, el servicio omite este valor. Si no desea limitar la cantidad de memoria que el servicio de integración de datos puede asignar, establezca esta propiedad en 0.</p> <p>Si el valor es mayor que 0, el servicio de integración de datos utiliza la propiedad para calcular la memoria total máxima permitida para ejecutar todas las solicitudes de forma simultánea. El servicio de integración de datos calcula la memoria total máxima de la siguiente forma:</p> <p>Tamaño máximo de memoria + Tamaño de heap máximo + Memoria necesaria para cargar componentes de programa</p> <p>El valor predeterminado es 0.</p> <p>Nota: Si ejecuta perfiles o asignaciones de calidad de datos, establezca esta propiedad en 0.</p>

En la siguiente tabla se describe la propiedad Tamaño de heap máximo para el proceso del Servicio de integración de datos:

Propiedad	Descripción
Tamaño de heap máximo	<p>Cantidad de RAM asignada a la máquina virtual de Java (JVM) que ejecuta el servicio de integración de datos. Esta propiedad se utiliza para aumentar el rendimiento. Añada una de las siguientes letras al valor para especificar las unidades:</p> <ul style="list-style-type: none">- b para bytes.- k para kilobytes.- m para megabytes.- g para gigabytes. <p>El valor predeterminado es 640 megabytes.</p> <p>Nota: Considere la posibilidad de aumentar el tamaño de heap máximo cuando el servicio de integración de datos necesite procesar grandes cantidades de datos.</p>

Memoria caché de objetos de datos

El servicio de integración de datos utiliza la memoria caché de objetos de datos para acceder a tablas virtuales y objetos de datos lógicos pregenerados. Habilite la memoria caché de objetos de datos para aumentar el rendimiento de las asignaciones, las consultas del servicio de datos SQL y las solicitudes de servicio web que incluyen objetos de datos lógicos y tablas virtuales.

De forma predeterminada, el servicio de integración de datos extrae los datos de origen y genera los objetos de datos requeridos cuando ejecuta una asignación, una consulta del servicio de datos SQL o una solicitud del servicio web. Cuando habilita la memoria caché de objetos de datos, el servicio de integración de datos puede utilizar los objetos de datos lógicos y las tablas virtuales almacenados en la memoria caché.

Realice los siguientes pasos para configurar la memoria caché de objetos de datos para objetos de datos lógicos y tablas virtuales en una aplicación:

1. Configure la conexión de base de datos de la memoria caché de objetos de datos en las propiedades de la memoria caché del servicio de integración de datos.
2. Habilite la memoria caché en las propiedades de objetos de datos lógicos o tablas virtuales en una aplicación.

De forma predeterminada, el componente Administrador de memoria caché de objetos de datos del servicio de integración de datos administra las tablas de la memoria caché para los objetos de datos lógicos y las tablas virtuales en la base de datos de la memoria caché de objetos de datos. Cuando el Administrador de memoria caché de objetos de datos administra la memoria caché, inserta todos los datos en las tablas de la memoria caché en cada actualización. Si desea actualizar las tablas de la memoria caché de forma incremental, puede administrar personalmente las tablas de la memoria caché mediante un cliente de base de datos u otra herramienta externa. Una vez habilitada la memoria caché de objetos de datos, puede configurar un objeto de datos lógicos o una tabla virtual para utilizar una tabla de memoria caché administrada por el usuario.

Para utilizar la marca de tiempo con el tipo de datos de zona horaria y habilitar el almacenamiento en la memoria caché de objetos de datos para IBM DB2 o Microsoft SQL Server, defina de la siguiente manera el formato de fecha y hora de la asignación implementada: "YYYY-MM-DD HH24:MI:SS". El servicio de integración de datos escribe los datos con una precisión de segundos.

Tipos de datos de tablas de la memoria caché

El Servicio de integración de datos utiliza datos de las tablas de la memoria caché cuando procesa asignaciones, consultas del servicio de datos SQL y solicitudes de servicio web que contienen objetos almacenados en la memoria caché. Los tipos de datos de las tablas de la memoria caché que el Servicio de integración de datos espera pueden diferir de los tipos de datos de los objetos almacenados en la memoria caché.

El Administrador de memoria caché de objetos de datos crea las tablas de la memoria caché con los tipos de datos que el Servicio de integración de datos espera. Si utiliza tablas de la memoria caché administradas por el usuario, compruebe que las tablas de la memoria caché utilizan los tipos de datos que el Servicio de integración de datos de espera.

Tipos de datos de la memoria caché de tabla virtual

La siguiente tabla muestra los tipos de datos de tabla de la memoria caché para tablas virtuales:

Tipo de datos de tabla virtual	IBM DB2	Microsoft SQL Server	Oracle
Char	Vargraphic Dbclob, para una precisión superior a 32672	Nvarchar Ntext, para una precisión superior a 4000	Nvarchar2 Nclob, para una precisión superior a 2000
Bigint	Bigint	Bigint	Número
Booleano	Entero	Int	Número
Fecha	Marca de tiempo	Datetime2	Marca de tiempo
Doble	Doble	Float	Marca de tiempo
Decimal	Decimal	Decimal	Número
Int	Entero	Int	Número
Hora	Marca de tiempo	Datetime2	Marca de tiempo
Marca de tiempo	Marca de tiempo	Datetime2	Marca de tiempo
Varbinary	Blob	Binario Imagen, para una precisión superior a 8000	Raw Blob, para una precisión superior a 2000
Varchar	Vargraphic Dbclob, para una precisión superior a 32672	Nvarchar Ntext, para una precisión superior a 4000	Nvarchar2 Nclob, para una precisión superior a 2000

Tipos de datos de la memoria caché de objetos de datos lógicos

La siguiente tabla muestra los tipos de datos de tabla de la memoria caché para objetos de datos lógicos:

Tipo de datos de objetos de datos lógicos	DB2	Microsoft SQL Server	Oracle
Bigint	Bigint	Bigint	Número
Binario	Blob	Binario Imagen, para una precisión superior a 8000	Raw Blob, para una precisión superior a 2000
Fecha/hora	Marca de tiempo	Datetime2	Marca de tiempo
Doble	Doble	Float	Número
Decimal	Decimal	Decimal	Número
Entero	Entero	Int	Número
String	Vargraphic Dbclob, para una precisión superior a 32672	Nvarchar Ntext, para una precisión superior a 4000	Nvarchar2 Nclob, para una precisión superior a 2000
Texto	Vargraphic Dbclob, para una precisión superior a 32672	Nvarchar Ntext, para una precisión superior a 4000	Nvarchar2 Nclob, para una precisión superior a 2000

Optimización de la memoria caché de objetos de datos

El rendimiento de la memoria caché depende del rendimiento de la base de datos de la memoria caché y de la configuración de los objetos en las asignaciones, los servicios de datos SQL y los servicios web.

Tenga en cuenta las siguientes soluciones para aumentar el rendimiento de la memoria caché:

Optimice la base de datos de la memoria caché.

El rendimiento óptimo de la memoria caché depende de la velocidad y del rendimiento de la base de datos de la memoria caché y del tamaño de la memoria caché. Configure el tamaño de la memoria caché dentro de la base de datos de la memoria caché.

Dado que el administrador de la memoria caché de objetos de datos debe mantener la memoria caché antigua durante una operación de actualización, la memoria caché debe ser lo suficientemente grande como para almacenar dos conjuntos de datos. Utilice la siguiente fórmula para calcular el tamaño mínimo requerido de la memoria caché:

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

Por ejemplo, desea almacenar en la memoria caché 20 objetos de datos lógicos y 10 tablas virtuales. Si el tamaño medio de los objetos es de 15 MB, entonces el tamaño requerido para la memoria caché será de $2 * 15 \text{ MB} * (20 + 10) = 900 \text{ MB}$.

Las tablas de la memoria caché son de solo lectura. Los usuarios finales no pueden actualizar las tablas de la memoria caché con los comandos SQL.

Defina claves principales y claves externas para los objetos de datos lógicos.

Cuando el Servicio de integración de datos genera memoria caché para objetos de datos lógicos con claves, crea índices. Los índices pueden aumentar el rendimiento de las consultas en la base de datos de la memoria caché.

Almacene en caché los objetos de datos lógicos que combina en una asignación.

Cuando combina objetos de datos lógicos almacenados en la memoria caché, el Servicio de integración de datos puede insertar la lógica de transformación de combinación en la base de datos de la memoria caché incluso cuando los datos de origen se originan en bases de datos diferentes.

Genere la memoria caché del índice en función de las columnas en un objeto de datos lógicos o una tabla virtual.

Configure el Servicio de integración de datos para generar una memoria caché del índice en función de las columnas en los objetos de datos lógicos o las tablas virtuales. El índice puede aumentar el rendimiento de las consultas en la base de datos de la memoria caché.

Optimización del sistema

Con frecuencia el rendimiento se reduce porque la asignación depende de conexiones ineficientes o debido a la sobrecarga del sistema de procesos del servicio de integración de datos. En otros casos los retrasos del sistema se deben a enrutadores, conmutadores, protocolos de red o el uso por muchos usuarios.

Reducir el acceso al disco en bases de datos de origen y destino, en sistemas de archivos de origen y destino y en los nodos del dominio puede reducir el rendimiento de la asignación. Pida al administrador del sistema que evalúe los discos duro de los equipos.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de la optimización del sistema:

Mejorar la velocidad de red.

Las conexiones de red lentas pueden reducir el rendimiento de la asignación. Pida al administrador del sistema que determine si la red se ejecuta a la velocidad óptima. Reduzca el número de saltos de red entre el proceso del servicio de integración de datos y las bases de datos.

Utilice varias CPU.

Puede usar varias CPU para ejecutar varias asignaciones en paralelo.

Reduzca la paginación.

Cuando un sistema operativo se ejecuta fuera de la memoria física, empieza paginando en el disco en la física libre. Configure la memoria física para el equipo del proceso del servicio de integración de datos para minimizar la paginación en el disco.

Utilice el enlace del procesador.

En un entorno UNIX de procesador múltiple, el servicio de integración de datos puede utilizar una gran cantidad de recursos del sistema. Utilice el enlace del procesador para controlar el uso del procesador mediante el proceso del servicio de integración. Asimismo, si las bases de datos de origen y destino están en el mismo equipo, utilice el enlace del procesador para limitar los recursos utilizados por la base de datos.

CAPÍTULO 8

Optimización del servicio de datos SQL

Este capítulo incluye los siguientes temas:

- [Resumen de la optimización del servicio de datos SQL, 70](#)
- [Optimización de la herramienta de cliente de terceros, 71](#)
- [Niveles del optimizador del servicio de datos SQL, 71](#)
- [Propiedades del servicio de datos SQL para la memoria y las solicitudes simultáneas, 75](#)
- [Memoria caché del conjunto de resultados para un servicio de datos SQL, 76](#)
- [Conservación de datos virtuales en tablas temporales, 78](#)

Resumen de la optimización del servicio de datos SQL

Puede optimizar los servicios de datos SQL para mejorar el rendimiento cuando los usuarios finales ejecutan consultas SQL en los mismos mediante herramientas de cliente de terceros. Si un servicio de datos SQL utiliza una asignación de tablas virtuales, puede optimizar los orígenes, las transformaciones y la asignación.

Utilice las siguientes técnicas de optimización para optimizar un servicio de datos SQL:

- Optimice las herramientas de cliente de terceros.
- Configure el nivel del optimizador del servicio de datos SQL.
- Configure las propiedades del servicio de datos SQL para la simultaneidad y memoria para un proceso de integración de datos.
- Configure la memoria caché de objetos de datos para el servicio de datos SQL.
- Configure la memoria caché del conjunto de resultados para el servicio de datos SQL.
- Configure las restricciones para tablas virtuales en el servicio de datos SQL.

TEMAS RELACIONADOS

- [“Memoria caché de objetos de datos” en la página 66](#)

Optimización de la herramienta de cliente de terceros

Las herramientas de cliente de terceros pueden afectar al rendimiento durante el procesamiento y la ejecución de consultas SQL frente a un servicio de datos SQL. Optimice las herramientas de cliente de terceros que los usuarios finales pueden utilizar para ejecutar consultas SQL frente a un servicio de datos SQL.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de la herramienta de cliente de terceros:

Envíe los resultados grandes de las consultas a un archivo en el disco.

Una herramienta de cliente de terceros puede afectar al rendimiento si muestra grandes resultados de consultas en la ventana de la consola.

Configure la herramienta de cliente de terceros para deshabilitar el cifrado.

Una herramienta de cliente de terceros puede afectar al rendimiento si cifra los datos al obtener o mostrar los resultados de las consultas.

Configure la herramienta de cliente de terceros para obtener un conjunto de filas previamente.

Una herramienta de cliente de terceros puede afectar al rendimiento si obtiene filas únicas al mismo tiempo.

Configure la herramienta de cliente de terceros para deshabilitar la opción para leer el contenido de la tabla cuando se ha cargado primero.

Una herramienta de cliente de terceros puede afectar al rendimiento si la configuración de tipos de datos para los tipos de datos BLOB y CLOB está establecida para leer el contenido de la tabla cuando se carga primero si los tipos de datos BLOB y CLOB no se utilizan en la consulta.

Configure la herramienta de cliente de terceros para utilizar el formato predeterminado y la configuración de conversión para la Fecha, Hora, y Marca de tiempo.

Una herramienta de cliente de terceros puede afectar al rendimiento si el formato de Fecha, Hora, y Marca de tiempo y la configuración de conversión están establecidos en un formato definido por el usuario en lugar de en el formato predeterminado.

Deshabilite la opción de depuración o establezca la misma en no depurar.

Una herramienta de cliente de terceros puede afectar al rendimiento si la opción de depuración para ejecutar la consulta está establecida en seguimiento. Esto puede reducir el rendimiento a medida que la herramienta de cliente de terceros escribe más mensajes de registro en el archivo de depuración durante el procesamiento de la consulta.

Niveles del optimizador del servicio de datos SQL

El servicio de integración de datos optimiza los servicios de datos SQL en función del nivel del optimizador que configure. Configure el nivel del optimizador cuando desee que el servicio de datos SQL use un nivel de

optimizador distinto del normal. De forma predeterminada, cada servicio de datos SQL utiliza el nivel del optimizador normal.

Para comprender cómo el nivel del optimizador crea una consulta optimizada para un servicio de datos SQL, vea el plan de consulta para un servicio de datos SQL. Cuando vea el plan de consulta, Developer Tool muestra una representación gráfica de la consulta optimizada en función del nivel del optimizador y una representación gráfica de la consulta original.

Puede configurar los siguientes niveles del optimizador:

0 (Ninguno)

El servicio de integración de datos no aplica ninguna optimización.

1 (Mínimo)

El servicio de integración de datos aplica el método de optimización de primera proyección.

2 (Normal)

El servicio de integración de datos aplica los métodos de optimización de primera proyección, de primera selección, de eliminación de ramificaciones, de inserción, de predicado global y de predicado.

3 (Completo)

El servicio de integración de datos aplica los métodos de optimización basada en el coste, de primera proyección, de primera selección, de eliminación de ramificaciones, de predicado, de inserción, semi-join y dataship-join.

El valor predeterminado es 2 (Normal).

Puede utilizar uno o varios de los siguientes métodos para configurar el nivel del optimizador para un servicio de datos SQL:

- Configure el nivel del optimizador para la vista previa de datos de los servicios de datos SQL.
- Configure el nivel de optimización de los servicios de datos SQL implementados.
- Configure el nivel del optimizador en la cadena de conexión de las consultas que se ejecutan sobre una implementación de servicios de datos SQL.

Configuración del nivel del optimizador del servicio de datos SQL para la vista previa de datos

Configure el nivel del optimizador que utiliza el servicio de integración de datos para ejecutar consultas SQL cuando se previsualiza la salida de un servicio de datos SQL.

1. En Developer Tool, haga clic en **Ejecutar > Abrir cuadro de diálogo Ejecutar**.
Aparece el cuadro de diálogo **Ejecutar**.
2. Haga clic en **Configuración del visor de datos**.
3. Haga clic en el botón **Nueva**.
4. Especifique un nombre para la configuración del visor de datos.
5. Haga clic en la ficha **Opciones avanzadas**.
6. Seleccione un nivel de optimizador.
7. Haga clic en **Aplicar**.
8. Haga clic en **Cerrar**.
Developer Tool crea la configuración del visor de datos.

Configuración del nivel del optimizador para los servicios de datos SQL implementados

Configure el nivel del optimizador que usan los servicios de integración de datos para ejecutar consultas SQL en un servicio de datos SQL implementado. Si desea reemplazar el nivel del optimizador para una sola consulta, puede configurarlo en la conexión del servicio de datos SQL.

1. En Administrator Tool, seleccione un servicio de integración de datos.
2. Haga clic en la vista **Aplicaciones**.
3. Expanda la aplicación que contiene el servicio de datos SQL para el que desea configurar el nivel del optimizador.
4. Seleccione el servicio de datos SQL y edite la siguiente propiedad:

Propiedad	Descripción
Nivel de optimización	<p>El nivel del optimizador que el Servicio de integración de datos aplica en el objeto. Introduzca el valor numérico que está asociado con el nivel del optimizador que desee configurar. Puede introducir uno de los siguientes valores numéricos:</p> <ul style="list-style-type: none">- 0. El Servicio de integración de datos no aplica la optimización.- 1. El Servicio de integración de datos aplica el método de optimización de primera proyección.- 2. El Servicio de integración de datos aplica los métodos de optimización de primera proyección, de primera selección, de inserción y de predicado.- 3. El Servicio de integración de datos aplica los métodos de optimización basados en coste de primera proyección, primera selección, inserción, predicado y semi-join.

5. Para reemplazar el nivel del optimizador que usan los servicios de integración de datos para ejecutar una consulta, añada la siguiente entrada a la URL de JDBC o la cadena de conexión de ODBC:

```
SQLDataServiceOptions.optimizeLevel= <nivel_optimizador_numérico>.
```

Plan de consulta del servicio de datos SQL

Cuando vea el plan de consulta de un servicio de datos SQL, puede ver una representación gráfica de la consulta original y la representación gráfica de la consulta optimizada. La representación gráfica describe cómo procesa la consulta el servicio de integración de datos. Incluye las transformaciones y el orden en que los servicios de integración de datos procesa cada transformación.

Developer Tool utiliza el nivel del optimizador que se establece en Developer Tool para generar la consulta optimizada. La consulta optimizada muestra la consulta tal como la ejecuta el servicio de integración de datos.

Por ejemplo, desea consultar la tabla virtual CUSTOMERS en un servicio de datos SQL. En la vista **Visor de datos**, elija los ajustes de configuración del visor de datos predeterminado para establecer el nivel del optimizador para la consulta como normal.

Introduzca la siguiente consulta en la vista **Visor de datos**:

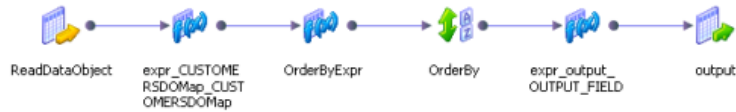
```
select * from CUSTOMERS where CUSTOMER_ID > 150000 order by LAST_NAME
```

Cuando vea el plan de consulta SQL, Developer Tool mostrará la siguiente representación gráfica de la consulta:



La vista no optimizada muestra la consulta que especifique. Developer Tool muestra la cláusula WHERE como una transformación de filtro y la cláusula ORDER BY como una transformación de ordenación. Developer Tool utiliza la transformación de expresión de exclusión para cambiar el nombre de los puertos.

Cuando vea la consulta optimizada, Developer Tool mostrará la siguiente representación gráfica de la consulta:



La vista optimizada muestra la consulta que ejecuta el servicio de integración de datos. Puesto que el nivel del optimizador es normal, el servicio de integración de datos inserta la condición de filtro en el objeto de datos de origen. El hecho de insertar la condición de filtro aumenta el rendimiento de la consulta, ya que reduce el número de filas que lee el servicio de integración de datos desde el objeto de datos de origen. Al igual que en la consulta no optimizada, Developer Tool muestra la cláusula ORDER BY como una transformación de ordenación. Utiliza transformaciones de expresión de exclusión para poner en práctica los tipos de datos especificados en las transformaciones lógicas.

Cómo ver un plan de consulta SQL

Muestre el plan de consulta SQL para ver una representación similar a una asignación de la consulta SQL que se introduce en el momento de previsualizar los datos de una tabla virtual.

1. Abra un servicio de datos SQL que contenga, como mínimo, una tabla virtual.
2. Haga clic en la vista **Visor de datos**.
3. Especifique una consulta SQL en la ventana **Entrada**.
4. Si lo desea, seleccione una configuración de visor de datos que contenga el nivel del optimizador que desee aplicar a la consulta.
5. Haga clic en **Mostrar plan de la consulta**.

Developer Tool muestra el plan de la consulta SQL para la consulta introducida en la ficha **No optimizada**.

6. Para ver la consulta optimizada, haga clic en la ficha **Optimizada**.

Developer Tool muestra el plan de la consulta SQL HypoStores_Customers.

Propiedades del servicio de datos SQL para la memoria y las solicitudes simultáneas

Para optimizar el rendimiento del servicio de datos SQL, configure las propiedades de simultaneidad y de memoria para el Servicio de integración de datos en la Herramienta del administrador.

En la siguiente tabla se describe la propiedad Memoria máxima por solicitud para el Módulo del servicio de SQL:

Propiedad	Descripción
Memoria máxima por solicitud	<p>El comportamiento de Memoria máxima por solicitud depende de las siguientes configuraciones del servicio de integración de datos:</p> <ul style="list-style-type: none">- El servicio ejecuta trabajos en procesos locales o remotos independientes, o la propiedad del servicio Tamaño máximo de memoria es 0 (predeterminado). <p>En este caso, Memoria máxima por solicitud es la cantidad máxima de memoria, en bytes, que el servicio de integración de datos puede asignar a todas las transformaciones que utilizan el modo de memoria caché automática en una sola solicitud. El servicio asigna la memoria por separado a las transformaciones que tienen un tamaño de memoria caché específico. La memoria total utilizada por la solicitud puede superar el valor de Memoria máxima por solicitud.</p> <ul style="list-style-type: none">- El servicio ejecuta trabajos en el proceso de servicio de integración de datos, y la propiedad del servicio Tamaño máximo de memoria es mayor que 0. <p>En este caso, Memoria máxima por solicitud es la cantidad máxima de memoria, en bytes, que el servicio de integración de datos puede asignar a una sola solicitud. La memoria total utilizada por la solicitud no puede superar el valor de Memoria máxima por solicitud.</p> <p>El valor predeterminado es 50 000 000.</p>

En la siguiente tabla se describe la propiedad Tamaño de heap máximo para el proceso del Servicio de integración de datos:

Propiedad	Descripción
Tamaño de heap máximo	<p>Cantidad de RAM asignada a la máquina virtual de Java (JVM) que ejecuta el servicio de integración de datos. Esta propiedad se utiliza para aumentar el rendimiento. Añada una de las siguientes letras al valor para especificar las unidades:</p> <ul style="list-style-type: none">- b para bytes.- k para kilobytes.- m para megabytes.- g para gigabytes. <p>El valor predeterminado es 640 megabytes.</p> <p>Nota: Considere la posibilidad de aumentar el tamaño de heap máximo cuando el servicio de integración de datos necesite procesar grandes cantidades de datos.</p>

En la siguiente tabla se describen las propiedades de SQL para el proceso del Servicio de integración de datos:

Propiedad	Descripción
Número máximo de conexiones simultáneas	Limita el número de conexiones de base de datos que puede llevar a cabo el Servicio de integración de datos para los servicios de datos SQL. El valor predeterminado es 100.

En la siguiente tabla se describen las opciones de ejecución para el Servicio de integración de datos:

Propiedad	Descripción
Tamaño máximo del grupo de ejecución a petición	Número máximo de trabajos a petición que pueden ejecutarse simultáneamente. Los trabajos incluyen previsualizaciones de datos, trabajos de creación de perfiles, consultas REST y SQL, solicitudes de servicios web y asignaciones que se ejecutan desde Developer tool. Todos los trabajos que recibe el servicio de integración de datos contribuyen al tamaño del grupo a petición. El servicio de integración de datos ejecuta inmediatamente trabajos a petición si hay suficientes recursos disponibles. De lo contrario, el servicio de integración de datos rechaza el trabajo. El valor predeterminado es 10.
Tamaño máximo del grupo de ejecución de lotes nativos	Número máximo de trabajos implementados que se pueden ejecutar simultáneamente en el entorno nativo. El servicio de integración de datos mueve trabajos nativos de asignación desde la cola hasta el grupo de trabajos nativos cuando hay suficientes recursos disponibles. El valor predeterminado es 10.
Tamaño máximo del grupo de ejecución de lotes Hadoop	Número máximo de trabajos implementados que se pueden ejecutar simultáneamente en el entorno de Hadoop. El servicio de integración de datos mueve trabajos de Hadoop desde la cola hasta el grupo de lotes de Hadoop cuando hay suficientes recursos disponibles. El valor predeterminado es 100.
Tamaño máximo de memoria	<p>Cantidad máxima de memoria, en bytes, que el servicio de integración de datos puede asignar para ejecutar todas las solicitudes de forma simultánea cuando el servicio ejecuta trabajos en el proceso de servicio de integración de datos. Cuando el servicio de integración de datos ejecuta trabajos en procesos locales o remotos independientes, el servicio omite este valor. Si no desea limitar la cantidad de memoria que el servicio de integración de datos puede asignar, establezca esta propiedad en 0.</p> <p>Si el valor es mayor que 0, el servicio de integración de datos utiliza la propiedad para calcular la memoria total máxima permitida para ejecutar todas las solicitudes de forma simultánea. El servicio de integración de datos calcula la memoria total máxima de la siguiente forma:</p> <p>Tamaño máximo de memoria + Tamaño de heap máximo + Memoria necesaria para cargar componentes de programa</p> <p>El valor predeterminado es 0.</p> <p>Nota: Si ejecuta perfiles o asignaciones de calidad de datos, establezca esta propiedad en 0.</p>

Memoria caché del conjunto de resultados para un servicio de datos SQL

Cuando configure la memoria caché del conjunto de resultados, el servicio de integración de datos guarda en la memoria caché los resultados del proceso DTM asociados a cada consulta de servicio de datos SQL y cada solicitud de servicio web. El servicio de integración de datos almacena en la memoria caché los resultados durante el período de vencimiento que configure. Cuando un cliente hace la misma consulta antes de que venza el período establecido, el servicio de integración de datos devolverá los resultados almacenados en la memoria caché.

Tenga en cuenta la siguiente solución para los cuellos de botella de la memoria caché del conjunto de resultados:

Configure la memoria caché del conjunto de resultados para un servicio de datos SQL.

La memoria caché del conjunto de resultados permite que el servicio de integración de datos utilice los resultados almacenados en la memoria caché para las consultas del servicio de datos SQL. Los usuarios que ejecuten consultas idénticas en un corto período de tiempo puede que quieran usar la

memoria caché del conjunto de resultados para reducir el tiempo de ejecución de las consultas idénticas.

Al habilitar el servicio de integración de datos para utilizar los resultados almacenados en memoria caché, aumenta el rendimiento del servicio de datos. Sin embargo, para mejorar más el tiempo de procesamiento del servicio de datos para consultas idénticas, asigne espacio suficiente para almacenar la memoria caché en la memoria. Cuando configure la cantidad de memoria caché para que sea igual o mayor de lo que es necesario para los resultados de la memoria caché, puede aumentar el rendimiento reduciendo la sobrecarga E/S del sistema. Cuando el servicio de integración de datos escribe archivos de caché en el disco, el tiempo de procesamiento del servicio de datos aumenta debido a una sobrecarga E/S del sistema.

Propiedades de la memoria caché del conjunto de resultados del servicio de datos SQL

Para aumentar el rendimiento, puede configurar las propiedades de la memoria caché del conjunto de resultados para un servicio de integración de datos. También puede configurar el número de milisegundos que tiene disponibles la memoria caché del conjunto de resultados para utilizar para un servicio de datos SQL.

La siguiente tabla describe las propiedades de la memoria caché del conjunto de resultados para el servicio de integración de datos:

Propiedad	Descripción
Prefijo de nombre de archivo	Prefijo de los nombres de todos los archivos de la memoria caché del conjunto de resultados almacenados en el disco. El valor predeterminado es RSCACHE.
Habilitar cifrado	Indica si los archivos de la memoria caché del conjunto de resultados están cifrados mediante AES de 128 bits. Los valores válidos son true o false. El valor predeterminado es true.

La siguiente tabla describe la propiedad que configura el número de milisegundos que la memoria caché de conjunto de resultados tiene disponibles para el servicio de datos SQL:

Propiedad	Descripción
Período de vencimiento de la memoria caché del conjunto de resultados	El número de milisegundos en el que la memoria caché del conjunto de resultados está disponible para poder usarla. Si se ha establecido en -1, la memoria caché no vence nunca. Si se ha establecido en 0, se deshabilita la memoria caché del conjunto de resultados. Los cambios efectuados en el período de vencimiento no se aplican a las memorias caché actuales. Si desea que todas las memorias caché usen el mismo período de vencimiento, purgue la memoria caché del conjunto de resultados tras cambiar el período de vencimiento. El valor predeterminado es 0.

Cómo habilitar la memoria caché del conjunto de resultados para un servicio de datos SQL

Para utilizar los resultados almacenados en la memoria caché para consultas del servicio de datos SQL idénticas, configure el servicio de integración de datos para que utilice la memoria caché del conjunto de resultados.

1. En Administrator Tool, seleccione un servicio de integración de datos.

2. Haga clic en la vista **Proceso** para configurar las propiedades de la memoria caché del conjunto de resultados.
3. Haga clic en la vista **Aplicación** y, a continuación, haga clic en el servicio de datos SQL para configurar la propiedad de vencimiento de la memoria caché del conjunto de resultados.

Conservación de datos virtuales en tablas temporales

Una tabla temporal es una tabla de una base de datos relacional que almacena datos temporales e intermedios. Las consultas complejas con frecuencia requieren almacenamiento para grandes cantidades de datos intermedios, como información de combinaciones. Al implementar tablas temporales, las herramientas de inteligencia empresarial pueden recuperar estos datos de la tabla temporal en lugar del servicio de datos SQL. De esta forma, aumenta el rendimiento.

Las tablas temporales también proporcionan una mayor seguridad de dos maneras. En primer lugar, solo el usuario de la sesión activa puede acceder a las tablas. Asimismo, las tablas persisten mientras una sesión está activa, y la base de datos elimina las tablas cuando la conexión se cierra.

Implementación de tablas temporales

Puede utilizar tablas temporales para mejorar el rendimiento de las consultas complejas y de gran tamaño. Las tablas temporales mejoran el rendimiento porque las consultas a las tablas temporales de una base de datos relacionales son más rápidas que las consultas sucesivas al servicio de datos SQL para un mismo conjunto de datos.

La implementación de tablas temporales para mejorar el rendimiento requiere la actuación de un administrador de Informática y un desarrollador de herramientas de inteligencia empresarial.

Primero, el administrador de Informática crea una conexión con la base de datos relacionales y configura el Servicio de integración de datos para utilizar dicha conexión.

A continuación, el desarrollador de una herramienta de inteligencia empresarial (por ejemplo, IBM Cognos o SAP Business Objects) crea una conexión entre la herramienta de inteligencia empresarial y el servicio de datos SQL de Informática. La conexión utiliza el controlador ODBC o JDBC de Informática.

Cuando estas conexiones están activas, la herramienta de inteligencia empresarial puede crear y utilizar tablas temporales para procesar grandes cantidades de datos intermedios.

CAPÍTULO 9

Optimización del servicio web

Este capítulo incluye los siguientes temas:

- [Resumen de optimización del servicio web, 79](#)
- [Optimizar solicitudes HTTP, 80](#)
- [Compresión de mensajes del servicio web, 80](#)
- [Nivel del optimizador del servicio web, 80](#)
- [Propiedades de los servicios web para la memoria y para solicitudes simultáneas , 82](#)
- [Propiedad del servicio web para configurar una instancia del DTM activa, 84](#)
- [Almacenamiento en memoria caché del conjunto de resultados de un servicio web, 85](#)
- [Administración del registro del servicio web, 85](#)

Resumen de optimización del servicio web

Puede optimizar los servicios web para mejorar el rendimiento cuando el servicio de integración de datos ejecuta solicitudes de servicio web. Ajuste el servicio de integración de datos para administrar la memoria y administrar solicitudes de servicio web simultáneas. Para mejorar el rendimiento del servicio web, utilice la compresión de mensajes del servicio web, optimice las solicitudes HTTP y configure el objeto de datos y la memoria caché del conjunto de resultados, y configure los niveles de registro de errores.

Utilice las siguientes técnicas de optimización para optimizar un servicio web:

- Optimice las solicitudes HTTP.
- Comprima los mensajes de servicio web.
- Configure el nivel del optimizador del servicio web.
- Configure las propiedades de los servicios web para los procesos de integración de datos de simultaneidad y memoria.
- Configure el servicio de integración de datos para mantener un proceso DTM activo de modo que pueda procesar más de una solicitud de servicio web.
- Configure la memoria caché de objetos de datos para el servicio web.
- Configure la memoria caché del conjunto de resultados para los servicios web.
- Configure los niveles de registro de errores de tiempo de ejecución de los servicios web.

TEMAS RELACIONADOS

- [“Memoria caché de objetos de datos” en la página 66](#)

Optimizar solicitudes HTTP

Optimice solicitudes HTTP para reducir el número de solicitudes en el servidor web.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de las solicitudes HTTP:

Reduzca el tiempo de espera del socket para el cliente del servicio web.

El tiempo de espera del socket establece la cantidad de tiempo que espera el cliente antes de agotar el tiempo de espera de la solicitud HTTP. El cliente del servicio web puede colgarse si el valor del tiempo de espera del socket es grande.

Compresión de mensajes del servicio web

Puede optimizar el rendimiento del servicio web comprimiendo grandes mensajes web que se pasan entre proveedores.

Tenga en cuenta la siguiente solución para los cuellos de botella de mensajes del servicio web:

Habilite la compresión de mensajes SOAP para un cliente del servicio web.

La compresión de mensajes SOAP habilita el servicio web para comprimir el servicio web y recibir mensajes del cliente de servicio web comprimidos. El servicio web puede aceptar un mensaje SOAP con compresión GZip desde un cliente del servicio web.

Cuando el servicio de integración de datos recibe la respuesta del servicio web, comprueba el encabezado HTTP de codificación de contenido en el mensaje SOAP y descodifica el mensaje.

Nivel del optimizador del servicio web

El servicio de integración de datos optimiza los servicios web en función del nivel del optimizador que configure. Configure el nivel del optimizador cuando desee que el servicio web utilice un nivel del optimizador distinto del normal. De forma predeterminada, cada servicio web utiliza el nivel del optimizador normal.

Puede elegir uno de los siguientes niveles del optimizador:

0 (Ninguno)

El servicio de integración de datos no aplica ninguna optimización.

1 (Mínimo)

El servicio de integración de datos aplica el método de optimización de primera proyección.

2 (Normal)

El servicio de integración de datos aplica los métodos de optimización de primera proyección, de primera selección, de eliminación de ramificaciones, de inserción, de predicado global y de predicado.

3 (Completo)

El servicio de integración de datos aplica los métodos de optimización basada en el coste, de primera proyección, de primera selección, de eliminación de ramificaciones, de predicado, de inserción, semi-join y dataship-join.

El valor predeterminado es 2 (Normal).

Puede utilizar uno o varios de los siguientes métodos para configurar el nivel del optimizador para un servicio web:

- Configure el nivel del optimizador para la vista previa de datos de un servicio web antes de implementarlo en un servicio de integración de datos.
- Configure el nivel de optimización para servicios web implementado que se ejecutan en un servicio de integración de datos.
- Configure el nivel del optimizador en el encabezado de la solicitud de servicio web para un servicio web implementado.

Configuración del nivel del optimizador del servicio web para la vista previa de datos

Configure el nivel del optimizador que utilizan los servicios de integración de datos para previsualizar la salida de un servicio web.

1. En Developer Tool, haga clic en **Ejecutar > Abrir cuadro de diálogo Ejecutar**.
Aparece el cuadro de diálogo **Ejecutar**.
2. Haga clic en **Configuración del servicio web**.
3. Haga clic en el botón **Nueva**.
4. Especifique un nombre para la configuración del servicio web.
5. Haga clic en la ficha **Opciones avanzadas**.
6. Seleccione un nivel de optimizador.
7. Haga clic en **Aplicar**.
8. Haga clic en **Cerrar**

Developer Tool crea la configuración del servicio web.

Cuando ejecute el visor de datos para previsualizar la salida de una asignación de operación, seleccione la configuración del servicio web que incluye el nivel del optimizador que desea utilizar.

Configuración del nivel del optimizador para servicios web implementados

Configure el nivel del optimizador que usan los servicios de integración de datos para ejecutar un servicio web implementado. Puede elegir reemplazar el nivel del optimizador para una sola solicitud mediante la configuración del nivel del optimizador en el encabezado HTTP de la solicitud SOAP del servicio web.

1. En Administrator Tool, seleccione un servicio de integración de datos.
2. Haga clic en la vista **Aplicaciones**.
3. Expanda la aplicación que contiene el servicio web para el que desea configurar el nivel del optimizador.

4. Seleccione el servicio web y edite la siguiente propiedad:

Propiedad	Descripción
Nivel de optimización	<p>El nivel del optimizador que el Servicio de integración de datos aplica en el objeto. Introduzca el valor numérico que está asociado con el nivel del optimizador que desee configurar. Puede introducir uno de los siguientes valores numéricos:</p> <ul style="list-style-type: none">- 0. El Servicio de integración de datos no aplica la optimización.- 1. El Servicio de integración de datos aplica el método de optimización de primera proyección.- 2. El Servicio de integración de datos aplica los métodos de optimización de primera proyección, de primera selección, de inserción y de predicado.- 3. El Servicio de integración de datos aplica los métodos de optimización basados en coste de primera proyección, primera selección, inserción, predicado y semi-join.

5. Para reemplazar el nivel de optimización del servicio web para una solicitud de servicio web, incluya la siguiente entrada en el encabezado HTTP de la solicitud SOAP del servicio web:

`WebServiceOptions.optimizeLevel= <nivel_optimizador_numérico>.`

Propiedades de los servicios web para la memoria y para solicitudes simultáneas

Para mejorar el rendimiento de los servicios web, configure las propiedades de simultaneidad y de memoria para el Servicio de integración de datos y para cada servicio web en la Herramienta del administrador.

En la siguiente tabla se describe la propiedad Memoria máxima por solicitud para el Módulo de servicios web:

Propiedad	Descripción
Memoria máxima por solicitud	<p>El comportamiento de Memoria máxima por solicitud depende de las siguientes configuraciones del servicio de integración de datos:</p> <ul style="list-style-type: none">- El servicio ejecuta trabajos en procesos locales o remotos independientes, o la propiedad del servicio Tamaño máximo de memoria es 0 (predeterminado). En este caso, Memoria máxima por solicitud es la cantidad máxima de memoria, en bytes, que el servicio de integración de datos puede asignar a todas las transformaciones que utilizan el modo de memoria caché automática en una sola solicitud. El servicio asigna la memoria por separado a las transformaciones que tienen un tamaño de memoria caché específico. La memoria total utilizada por la solicitud puede superar el valor de Memoria máxima por solicitud.- El servicio ejecuta trabajos en el proceso de servicio de integración de datos, y la propiedad del servicio Tamaño máximo de memoria es mayor que 0. En este caso, Memoria máxima por solicitud es la cantidad máxima de memoria, en bytes, que el servicio de integración de datos puede asignar a una sola solicitud. La memoria total utilizada por la solicitud no puede superar el valor de Memoria máxima por solicitud. El valor predeterminado es 50 000 000.

En la siguiente tabla se describen las opciones de ejecución para el Servicio de integración de datos:

Propiedad	Descripción
Tamaño máximo de memoria	<p>Cantidad máxima de memoria, en bytes, que el servicio de integración de datos puede asignar para ejecutar todas las solicitudes de forma simultánea cuando el servicio ejecuta trabajos en el proceso de servicio de integración de datos. Cuando el servicio de integración de datos ejecuta trabajos en procesos locales o remotos independientes, el servicio omite este valor. Si no desea limitar la cantidad de memoria que el servicio de integración de datos puede asignar, establezca esta propiedad en 0.</p> <p>Si el valor es mayor que 0, el servicio de integración de datos utiliza la propiedad para calcular la memoria total máxima permitida para ejecutar todas las solicitudes de forma simultánea. El servicio de integración de datos calcula la memoria total máxima de la siguiente forma:</p> <p>Tamaño máximo de memoria + Tamaño de heap máximo + Memoria necesaria para cargar componentes de programa</p> <p>El valor predeterminado es 0.</p> <p>Nota: Si ejecuta perfiles o asignaciones de calidad de datos, establezca esta propiedad en 0.</p>

En la siguiente tabla se describen las propiedades de configuración de HTTP para el proceso del Servicio de integración de datos:

Propiedad	Descripción
Número máximo de solicitudes acumuladas	Número máximo de conexiones HTTP o HTTPS que pueden estar en espera en una cola para este proceso de servicio de integración de datos. El valor predeterminado es 100.
Número máximo de solicitudes simultáneas	<p>Número máximo de conexiones HTTP o HTTPS que se pueden establecer con este proceso de servicio de integración de datos. El valor mínimo es 4. El valor predeterminado es 200.</p> <p>Nota: Para un servicio web, esta propiedad repercute en el número de solicitudes del servicio web que los servicios de integración de datos aceptan antes de enviar las solicitudes a las acumuladas en el Servicio de integración de datos.</p>

En la siguiente tabla se describe la propiedad Tamaño de heap máximo que puede configurar para el proceso del Servicio de integración de datos:

Propiedad	Descripción
Tamaño de heap máximo	<p>Cantidad de RAM asignada a la máquina virtual de Java (JVM) que ejecuta el servicio de integración de datos. Esta propiedad se utiliza para aumentar el rendimiento. Añada una de las siguientes letras al valor para especificar las unidades:</p> <ul style="list-style-type: none"> - b para bytes. - k para kilobytes. - m para megabytes. - g para gigabytes. <p>El valor predeterminado es 640 megabytes.</p> <p>Nota: Considere la posibilidad de aumentar el tamaño de heap máximo cuando el servicio de integración de datos necesite procesar grandes cantidades de datos.</p>

Ejemplo de configuración del Servicio de integración de datos para solicitudes de servicio web simultáneas

Al configurar cómo el Servicio de integración de datos procesa solicitudes de servicios web simultáneas, compruebe que el valor para el número máximo de solicitudes simultáneas es el mismo para proceso del servicio web y del Servicio de integración de datos.

Por ejemplo, en la siguiente configuración el Servicio de integración de datos acepta 200 solicitudes HTTP simultáneas, pero sólo 10 son solicitudes simultáneas del servicio web:

Tipo de propiedad	Nombre de la propiedad	Configuración
Proceso del Servicio de integración de datos	Número máximo de solicitudes simultáneas	200
Proceso del Servicio de integración de datos	Número máximo de solicitudes acumuladas	500
Servicio de integración de datos	Tamaño máximo del grupo de ejecución a petición	100
Servicio web	Número máximo de solicitud simultánea	10

Cuando el Servicio de integración de datos recibe 20 solicitudes de servicio web, 10 solicitudes de servicio web fallan porque el servicio web solo puede aceptar 10 solicitudes simultáneas.

Para evitar que las solicitudes de servicio web fallen cuando el servicio web alcanza el número máximo de solicitudes simultáneas, configure el mismo valor máximo para el proceso del Servicio de integración de datos y el servicio web. Cuando el número de solicitudes enviadas al Servicio de integración de datos excede el número máximo de solicitudes simultáneas, las solicitudes adicionales permanecen acumuladas hasta que el proceso del Servicio de integración de datos está disponible para procesar las solicitudes.

Propiedad del servicio web para configurar una instancia del DTM activa

Para aumentar el rendimiento, puede configurar el servicio de integración de datos para mantener una instancia del DTM activa de modo que pueda procesar más de una solicitud de servicio web. Puede configurar la propiedad tiempo para mantener DTM para el servicio de integración de datos en Administrator Tool.

La siguiente tabla describe la propiedad tiempo para mantener DTM:

Propiedad	Descripción
Tiempo para mantener DTM	<p>Número de milisegundos que la instancia de DTM permanece abierta tras completar la última solicitud. Las solicitudes de servicio web que se han emitido en la misma operación pueden reutilizar la instancia abierta. Utilice el tiempo para mantener para aumentar el rendimiento cuando el tiempo necesario para procesar la solicitud sea corto en comparación con el tiempo de inicialización para la instancia de DTM. Si la solicitud genera un error, la instancia de DTM finaliza. El valor predeterminado es 5000.</p> <p>Nota: La capacidad de utilizar una instancia del DTM existente aumenta el rendimiento. DIS requiere recursos adicionales para iniciar una instancia del DTM para cada solicitud. Mantener el DTM activo consume memoria. Por tanto, los usuarios deben tener en cuenta el consumo de memoria al configurar esta opción.</p>

Almacenamiento en memoria caché del conjunto de resultados de un servicio web

Cuando se configura el almacenamiento en memoria caché del conjunto de resultados, el servicio de integración de datos almacena en la memoria caché los resultados del proceso DTM asociado con cada una de las solicitudes de servicio web. El servicio de integración de datos almacena en la memoria caché los resultados durante el período de vencimiento que configure. Si un cliente externo repite la misma solicitud antes de que venza el período establecido, el servicio de integración de datos devolverá los resultados guardados en la memoria caché.

Tenga en cuenta la siguiente solución para los cuellos de botella de la memoria caché del conjunto de resultados:

Configure la memoria caché del conjunto de resultados para un servicio web.

El almacenamiento en memoria caché del conjunto de resultados permite que el servicio de integración de datos utilice los resultados almacenados en memoria caché para las solicitudes de un servicio web. Para usuarios que ejecuten consultas idénticas dentro de un breve espacio de tiempo, puede resultar beneficioso utilizar el almacenamiento en la memoria caché de conjunto de resultados para que las consultas sean más rápidas.

El servicio de integración de datos almacena en la memoria caché el conjunto de resultados para los servicios web por usuario cuando el servicio web utiliza la seguridad WS. El servicio de integración de datos almacena la memoria caché por el nombre de usuario proporcionado en el token username de la solicitud de servicio web. Cuando el servicio de integración de datos almacena en memoria caché los resultados por usuario, solamente devuelve los resultados almacenados en memoria caché al usuario que ha enviado la solicitud de servicio web.

Cómo habilitar la memoria caché del conjunto de resultados para un servicio web

Utilice los resultados almacenados en la memoria caché para solicitudes del servicio web idénticas, configure el servicio de integración de datos para utilizar la memoria caché del conjunto de resultados.

1. En Administrator Tool, seleccione un servicio de integración de datos.
2. Haga clic en la vista **Proceso** para configurar las propiedades de la memoria caché del conjunto de resultados.
3. Haga clic en la vista **Aplicación**, haga clic en el servicio web y, a continuación, haga clic en la operación para configurar el período de vencimiento de la memoria caché en las propiedades de operación del servicio web. Si desea que el servicio de integración de datos almacene en caché los resultados por usuario, habilite WS-Security en las propiedades del servicio web.
4. Para deshabilitar el almacenamiento en memoria caché del conjunto de resultados para una solicitud de servicio web cuando la operación del servicio web está configurada para almacenar en memoria caché el conjunto de resultados, incluya la siguiente sintaxis en el encabezado HTTP de la solicitud SOAP:

```
WebServiceOptions.disableResultSetCache=true
```

Administración del registro del servicio web

El rendimiento E/S del sistema puede reducirse cuando el servicio de integración de datos escribe y mantiene un gran número de archivos de registro. El servicio de integración de datos genera registros de

tiempo de ejecución del servicio web basados en el nivel de seguimiento que configure. Considere administrar el número de archivos de registro que el servicio de integración de datos escribe y mantiene.

Tenga en cuenta las siguientes soluciones los cuellos de botella del servicio web:

Desactive el nivel de seguimiento del servicio web.

Cuando configure las propiedades del servicio web para un servicio web implementado, puede especificar el nivel de seguimiento del registro. El nivel de seguimiento determina los tipos de registros que el servicio de integración de datos escribe en la ubicación del registro de tiempo de ejecución. El nivel de seguimiento del servicio web determinado es INFO. Cuando el nivel de seguimiento está establecido como FINEST o ALL, el rendimiento puede reducirse porque el servicio de integración de datos escribe más registros en el archivo de registro. El impacto en el rendimiento al establecer el nivel de seguimiento en FINEST o ALL es el mayor cuando el servicio web utiliza seguridad HTTPS y WS.

Archive los de archivos de registro que ya no son necesarios.

El sistema E/S se ve afectado al almacenar demasiados archivos de registro. De forma predeterminada, el servicio de integración de datos escribe los registros de tiempo de ejecución del servicio web en el siguiente directorio: <DirInstalaciónInformatica>/tomcat/bin/disLogs/ws

Nota: Si elimina la carpeta WS para vaciar los registros, debe volver a crearla. Detenga el servicio de integración de datos antes de eliminar y volver a crear la carpeta de WS.

Habilitar la propiedad para Omitir registros en el servicio de integración de datos

Active la propiedad **Omitir registros** para impedir que el servicio de integración de datos genere archivos de registro cuando la solicitud de servicio web se complete satisfactoriamente. El nivel de seguimiento de los servicios web debe establecerse en INFO o superior en el servicio de integración de datos.

CAPÍTULO 10

Optimización de las conexiones

Este capítulo incluye los siguientes temas:

- [Resumen de la optimización de las conexiones, 87](#)
- [Agrupación de conexiones, 87](#)
- [Tamaño del paquete de red de la base de datos, 89](#)

Resumen de la optimización de las conexiones

Puede optimizar las conexiones para mejorar el rendimiento. Puede administrar el grupo de instancias de conexiones inactivas para una conexión de base de datos. Puede aumentar el tamaño de paquetes de red para permitir que paquetes de mayor tamaño puedan cruzar la red al mismo tiempo.

Utilice las siguientes técnicas para optimizar las conexiones:

- Optimice la agrupación de conexiones.
- Optimice el tamaño del paquete de red de la base de datos.

Agrupación de conexiones

La agrupación de conexiones es un marco para almacenar en caché la información de conexión de base de datos usada por el servicio de integración de datos. Esto aumenta el rendimiento mediante la reutilización de la información de conexión almacenada en la memoria caché.

Tenga en cuenta la siguiente solución para los cuellos de botella de conexiones:

Habilite la agrupación de conexiones para una conexión de base de datos.

Habilite la agrupación de conexiones para optimizar el rendimiento de la conexión. Puede administrar las instancias de conexión inactivas para una conexión de base de datos. El grupo de conexiones conserva las instancias de conexión inactivas según las propiedades de agrupación configuradas. Puede ajustar el número máximo y mínimo de conexiones inactivas y el tiempo de espera máximo para una conexión inactiva.

Propiedades de agrupación en objetos de conexión

Puede editar las propiedades de agrupación de conexiones en la vista **Agrupación** para una conexión de base de datos.

El número de bibliotecas de grupos de conexiones depende del número de procesos del Servicio de integración de datos o de procesos DTM en ejecución. Cada proceso del Servicio de integración de datos o proceso DTM mantiene su propia biblioteca de grupos de conexiones. Los valores de las propiedades de agrupación son para cada biblioteca de grupos de conexiones.

Por ejemplo, si se establece el número máximo de conexiones en 15, cada biblioteca de grupos de conexiones puede tener un máximo de 15 conexiones inactivas en el grupo. Si el Servicio de integración de datos ejecuta tareas en procesos locales independientes y hay trse procesos DTM en ejecución, puede tener un máximo de 45 instancias de conexión inactiva.

Para reducir el número total de instancias de conexión inactivas, establezca el número mínimo de conexiones en 0 y reduzca el tiempo máximo de inactividad para cada conexión de base de datos.

En la siguiente tabla se describen las propiedades de agrupación de conexiones de base de datos que se pueden editar en la vista **Agrupación** para una conexión de base de datos

Habilitar agrupación de conexiones

Habilita la agrupación de conexiones. Al habilitar la agrupación de conexiones, cada grupo de conexiones guarda en memoria las instancias de conexión inactivas. Para eliminar los grupos de conexiones inactivas, se debe reiniciar el Servicio de integración de datos.

Si la agrupación de conexiones está deshabilitada, el proceso DTM o el proceso del Servicio de integración de datos detiene toda actividad de agrupación. El proceso DTM o el proceso del Servicio de integración de datos crea una instancia de conexión cada vez que procesa una tarea. Al finalizar el procesamiento de la tarea, descarta la instancia.

El valor predeterminado está habilitado para DB2 para i5/OS, DB2 para z/OS, IBM DB2, Microsoft SQL Server, Oracle y conexiones ODBC. La opción predeterminada está deshabilitada para las conexiones Adabas, IMS, Secuenciales y VSAM.

Número mínimo de conexiones

Número mínimo de instancias de conexión inactivas que un grupo mantiene para una conexión de base de datos después de que se alcance el tiempo máximo de inactividad. Este valor debe ser igual o inferior al número máximo de instancias de conexión inactivas. El valor predeterminado es 0.

Número máximo de conexiones

Número máximo de instancias de conexión inactivas que un grupo mantiene para una conexión de base de datos antes de que se alcance el tiempo máximo de inactividad. Este valor debe ser un número mayor que la cantidad mínima de instancias de conexión inactivas. El valor predeterminado es 15.

Tiempo máximo de inactividad

Número de segundos que una instancia de conexión que supera el número mínimo de instancias de conexión puede permanecer inactiva antes de que el grupo de conexiones la descarte. El grupo de conexiones omite el tiempo de inactividad cuando la instancia de conexión no supera el número mínimo de instancias de conexión inactivas. El valor predeterminado es 120.

Tamaño del paquete de red de la base de datos

Si lee o escribe en destinos de Oracle, ASE Sybase, o Microsoft SQL Server, puede mejorar el rendimiento incrementando el tamaño del paquete de red basado en la base de datos desde la que lee o en la que escribe. Aumentar el tamaño de paquetes de red permite que paquetes de datos de mayor tamaño puedan cruzar la red al mismo tiempo.

Tenga en cuenta las siguientes soluciones para los cuellos de botella de tamaño de los paquetes de red de la base de datos:

Aumente el tamaño de paquetes de red de base de datos para una base de datos Oracle.

Puede aumentar el tamaño del paquete de red del servidor de la base de datos en `listener.ora` y `tnsnames.ora`. Consulte la documentación de la base de datos para obtener más información sobre cómo aumentar el tamaño del paquete, si es necesario.

Aumente el tamaño de paquetes de red de base de datos para una base de datos ASE Sybase.

Consulte la documentación de la base de datos para obtener información sobre cómo aumentar el tamaño del paquete. También debe cambiar el tamaño del paquete para ASE Sybase en el objeto de conexión relacional en el servicio de integración de datos para que refleje el tamaño del paquete del servidor de la base de datos.

Aumente el tamaño de paquetes de red de base de datos para una base de datos de Microsoft SQL Server.

Consulte la documentación de la base de datos para obtener información sobre cómo aumentar el tamaño del paquete. También debe cambiar el tamaño del paquete para Microsoft SQL Server en el objeto de conexión relacional en el servicio de integración de datos para que refleje el tamaño del paquete del servidor de la base de datos.

INDICE

A

- Administración del registro del servicio web
 - nivel de seguimiento de error [86](#)
- agrupación de conexiones
 - optimización de las conexiones [87](#)
 - propiedades [88](#)
- ajuste del rendimiento
 - método de la optimización de inserción [49](#)
 - método de optimización basado en el coste [45](#)
 - método de optimización dataship-join [46](#)
 - método de optimización de eliminación de ramificaciones [49](#)
 - método de optimización de inserciones [49](#)
 - método de optimización de predicado [44](#)
 - método de optimización de predicado global [48](#)
 - método de optimización de primera proyección [43](#)
 - método de optimización de primera selección [48](#)
 - método de optimización semi-join [47](#)
 - métodos de optimización [42](#)
 - niveles de optimización [42](#)
 - resumen de proceso [10](#)
- archivos sin formato
 - optimización de destinos para la partición [57](#)
 - optimización de orígenes para la partición [57](#)
- asignación de memoria
 - Instancia del DTM activa [84](#)
 - Servicio de datos SQL [75](#)
 - Servicio web [84](#)
 - solicitudes simultáneas [75](#)
- asignaciones
 - método de optimización de predicado [44](#)
 - método de optimización de predicado global [48](#)
 - métodos de optimización [42](#)
 - optimización de particiones [55](#)

B

- bases de datos
 - optimización de destinos para la partición [58](#)
 - optimización de orígenes para la partición [58](#)

C

- cargas masivas
 - optimización de destino [15](#)
- cómo habilitar la memoria caché del conjunto de resultados para un servicio de datos SQL
 - memoria caché de conjunto de resultados [77](#)
- Compresión de mensajes del servicio web
 - Optimización del servicio web [80](#)
- Controladores de JDBC
 - optimización del tiempo de ejecución [72](#)
- creación de particiones
 - optimización [55](#)

- cuellos de botella
 - en UNIX [12](#)
 - en Windows [11](#)

D

- destino del archivo sin formato
 - optimización de destino [14](#)

E

- efectos secundarios
 - descripción [37](#)
 - Transformación de consumidor de servicio web [38](#)
 - transformación de SQL [34](#)
- eliminación del error de transformación
 - optimización de transformación [36](#)
- error de seguimiento
 - optimización de la asignación [53](#)
- Estadísticas de supervisión
 - optimización del tiempo de ejecución [64](#)

F

- filtros condicionales
 - optimización de origen [19](#)

H

- herramientas de cliente de terceros
 - optimización del tiempo de ejecución [71](#)

I

- Instancia del DTM activa
 - Servicio web [84](#)
- Intervalos de puntos de comprobación de base de datos
 - optimización de destino [15](#)

L

- lectura de pase individual
 - optimización de la asignación [52](#)

M

- memoria caché de conjunto de resultados
 - cómo habilitar la memoria caché del conjunto de resultados para un servicio de datos SQL [77](#)

- memoria caché de conjunto de resultados (*continuado*)
 - propiedades de memoria caché de conjunto de resultados [77](#)
- memoria caché de la transformación
 - optimización de transformación [35](#)
- memoria caché de objetos de datos
 - configurar [66](#)
 - descripción [66](#)
 - memoria caché de índice [66](#)
 - optimización [68](#)
 - tablas administradas por el usuario [66](#)
 - tipos de datos de tabla [67](#)
- Memoria caché del conjunto de resultados del servicio de datos SQL
 - Servicio de integración de datos [76](#)
- memoria caché del conjunto de resultados del servicio web
 - Servicio de integración de datos [85](#)
- método de optimización de inserción
 - inserción de origen [51](#)
- método de optimización de inserciones
 - inserción completa [50](#)

N

- nivel de optimización completa
 - descripción [42](#)
- nivel de optimización mínima
 - descripción [42](#)
- nivel de optimización normal
 - descripción [42](#)
- nivel de seguimiento de error
 - Administración del registro del servicio web [86](#)
- niveles de optimización
 - descripción [42](#)
- número máximo de paralelismos
 - cómo aumentar [56](#)

O

- objeto de datos personalizado
 - optimización de origen [22](#)
- objetos de datos lógicos
 - memoria caché en base de datos [66](#)
- optimización
 - efectos secundarios [37](#)
 - método de la optimización de inserción [49](#)
 - método de optimización basado en el coste [45](#)
 - método de optimización dataship-join [46](#)
 - método de optimización de eliminación de ramificaciones [49](#)
 - método de optimización de inserciones [49](#)
 - método de optimización de primera proyección [43](#)
 - método de optimización de primera selección [48](#)
 - método de optimización semi-join [47](#)
 - métodos de rendimiento de la asignación [42](#)
- optimización basada en el coste
 - descripción [45](#)
- optimización dataship-join
 - descripción [46](#)
- optimización de conversión de tipos de datos
 - optimización de la asignación [53](#)
- optimización de destino
 - cargas masivas [15](#)
 - destino del archivo sin formato [14](#)
 - Intervalos de puntos de comprobación de base de datos [15](#)
 - Optimización de la base de datos de Oracle [16](#)
- optimización de eliminación de ramificaciones
 - descripción [49](#)

- optimización de inserción
 - descripción [49](#)
 - habilitar en la transformación de SQL [35](#)
 - Transformación de consumidor de servicio web [39](#)
 - transformación de SQL [34](#)
- optimización de inserciones
 - descripción [49](#)
- optimización de la asignación
 - error de seguimiento [53](#)
 - lectura de pase individual [52](#)
 - optimización de conversión de tipos de datos [53](#)
 - optimización de la expresión [25](#)
 - optimización del filtro [52](#)
- Optimización de la base de datos de Oracle
 - optimización de destino [16](#)
 - optimización de origen [23](#)
- optimización de la consulta
 - optimización de origen [18](#)
- optimización de la expresión
 - optimización de la asignación [25](#)
- optimización de las conexiones
 - agrupación de conexiones [87](#)
 - tamaño del paquete de red de la base de datos [89](#)
- optimización de origen
 - filtros condicionales [19](#)
 - objeto de datos personalizado [22](#)
- Optimización de la base de datos de Oracle [23](#)
- optimización de la consulta [18](#)
- origen del archivo sin formato [18](#)
- restricciones [21](#)
- seleccionar distinto [19](#)
- optimización de primera proyección
 - descripción [43](#)
- optimización de primera selección
 - descripción [48](#)
 - Transformación de consumidor de servicio web [38](#)
 - transformación de SQL [34](#)
- optimización de transformación
 - eliminación del error de transformación [36](#)
 - memoria caché de la transformación [35](#)
 - transformación de agregación [24](#)
 - transformación de búsqueda [30](#)
 - Transformación de combinación [30](#)
 - Transformación de consumidor de servicio web [37](#)
 - transformación de Java [27](#)
 - Transformación de ordenación [33](#)
 - transformación de SQL [34](#)
- optimización del filtro
 - optimización de la asignación [52](#)
- Optimización del servicio de datos SQL
 - Controladores de JDBC [72](#)
 - herramientas de cliente de terceros [71](#)
- Optimización del servicio de integración de datos
 - optimización del tiempo de ejecución [62](#)
- Optimización del servicio de repositorio de modelos
 - optimización del tiempo de ejecución [63](#)
- Optimización del servicio del analista
 - optimización del tiempo de ejecución [62](#)
- Optimización del servicio web
 - Compresión de mensajes del servicio web [80](#)
 - optimizar solicitudes HTTP [80](#)
- optimización del sistema
 - optimización del tiempo de ejecución [69](#)
- optimización del tiempo de ejecución
 - Estadísticas de supervisión [64](#)
 - Optimización del servicio de integración de datos [62](#)
 - Optimización del servicio de repositorio de modelos [63](#)
 - Optimización del servicio del analista [62](#)

- optimización del tiempo de ejecución (*continuado*)
 - optimización del sistema [69](#)
- optimización semi-join
 - descripción [47](#)
- optimizar solicitudes HTTP
 - Optimización del servicio web [80](#)
- origen del archivo sin formato
 - optimización de origen [18](#)

P

- particiones
 - optimización de bases de datos de destino [58](#)
 - optimización de bases de datos de origen [58](#)
 - optimización de destinos de archivos sin formato [57](#)
 - optimización de orígenes de archivos sin formato [57](#)
 - optimización de transformaciones [59](#)
 - varias CPU [56](#)
- planes de consulta SQL
 - cómo ver [74](#)
- propiedades de memoria caché de conjunto de resultados
 - optimización del tiempo de ejecución [77](#)
- puerto de filtro
 - Transformación de consumidor de servicio web [39](#)

R

- restricciones
 - configurar restricciones [22](#)
 - optimización de origen [21](#)

S

- seleccionar distinto
 - optimización de origen [19](#)
- Servicio de datos SQL
 - asignación de memoria [75](#)
- Servicio de integración de datos
 - Memoria caché del conjunto de resultados del servicio de datos SQL [76](#)
 - memoria caché del conjunto de resultados del servicio web [85](#)
- Servicio web
 - asignación de memoria [84](#)
 - solicitudes simultáneas [82](#)
- sistema
 - cuellos de botella en UNIX, identificación [12](#)
 - cuellos de botella en Windows, identificación [11](#)
- solicitudes simultáneas
 - Servicio de datos SQL [75](#)
 - Servicio web [82](#)
- sugerencias
 - Vista de la consulta [21](#)
- sugerencias de la base de datos
 - presentación de Developer Tool [21](#)

- Sugerencias de SQL
 - presentación de Developer Tool [21](#)

T

- tablas temporales
 - descripción [78](#)
- tablas virtuales
 - memoria caché en base de datos [66](#)
- tamaño del paquete de red de la base de datos
 - optimización de las conexiones [89](#)
- Tiene efectos secundarios
 - descripción de la propiedad de la transformación [37](#)
- transformación de agregación
 - optimización de transformación [24](#)
- transformación de búsqueda
 - optimización de transformación [30](#)
- Transformación de combinación
 - optimización de transformación [30](#)
- Transformación de consumidor de servicio web
 - habilitar la optimización de inserción [40](#)
 - optimización de inserción [39](#)
 - optimización de primera selección [38](#)
 - optimización del filtro [39](#)
- transformación de Java
 - optimización de transformación [27](#)
- Transformación de ordenación
 - optimización de transformación [33](#)
- transformación de SQL
 - optimización de inserción [34](#)
 - optimización de primera selección [34](#)
 - optimización de transformación [34](#)
 - propiedades de la optimización de inserción [35](#)
- Transformación del consumidor de servicio web
 - optimización de transformación [37](#)
- transformaciones
 - optimización para partición [59](#)

U

- UNIX
 - cuellos de botella del sistema [12](#)

V

- Vista de la consulta
 - configurar sugerencias [21](#)

W

- Windows
 - cuellos de botella [11](#)