



Informatica®

10.1.1 HotFix 2

Guia de Ajuste de Desempenho

Este software e a documentação são fornecidos somente sob um contrato de licença separado, contendo restrições sobre uso e divulgação. Não está permitida de forma alguma a reprodução ou a transmissão de qualquer parte deste documento (seja por meio eletrônico, fotocópia, gravação ou quaisquer outros meios) sem o consentimento prévio da Informatica LLC.

Informatica e o logotipo Informatica são marcas comerciais ou marcas registradas da Informatica LLC nos Estados Unidos e em muitas jurisdições por todo o mundo. Uma lista atual das marcas comerciais da Informatica está disponível na Internet em <https://www.informatica.com/trademarks.html>. Os nomes de outras companhias e produtos podem ser nomes ou marcas comerciais de seus respectivos proprietários.

DIREITOS DO GOVERNO DOS ESTADOS UNIDOS Programas, softwares, bancos de dados, bem como a documentação e os dados técnicos relacionados, distribuídos a clientes do Governo dos EUA são "softwares de computador comerciais" ou "dados técnicos comerciais", de acordo com o Regulamento de Aquisição Federal aplicável e os regulamentos suplementares específicos da agência. Como tal, a utilização, duplicação, divulgação, modificação e adaptação estão sujeitas às restrições e aos termos de licença estabelecidos no contrato governamental aplicável e, na medida do que for aplicável pelos termos do contrato governamental, aos direitos adicionais estabelecidos no FAR 52.227-19, Licença de Software de Computador Comercial.

Partes deste software e/ou documentação estão sujeitas a copyright detido por terceiros. Os avisos de terceiros necessários são incluídos no produto.

As informações contidas neste documento estão sujeitas a alteração sem aviso prévio. Se você encontrar quaisquer problemas nesta documentação, informe-os em infa_documentation@informatica.com.

Os produtos Informatica apresentam garantias segundo os termos e condições dos acordos em que são fornecidos. A INFORMATICA FORNECE AS INFORMAÇÕES NESTE DOCUMENTO "COMO ESTÃO" SEM GARANTIA DE QUALQUER TIPO, EXPRESSA OU IMPLÍCITA, INCLUINDO, SEM QUAISQUER GARANTIAS DE COMERCIALIZAÇÃO, ADEQUAÇÃO A UM DETERMINADO FIM E QUALQUER GARANTIA OU CONDIÇÃO DE NÃO-VIOLAÇÃO.

Data da Publicação: 2018-07-23

Conteúdo

Prefácio.....	7
Recursos da Informatica.	7
Rede da Informatica.	7
Base de Dados de Conhecimento da Informatica.	7
Documentação da Informatica.	8
Matrizes de Disponibilidade de Produto Informatica.	8
Informatica Velocity.	8
Informatica Marketplace.	8
Suporte global a clientes Informatica.	8
 Capítulo 1: Visão Geral do Ajuste do Desempenho.....	9
Visão Geral do Ajuste do Desempenho.	9
Afunilamentos de Destinos.	10
Afunilamentos de Origens.	10
Afunilamentos de Mapeamentos.	11
Afunilamentos do Sistema de Computador.	11
Identificando afunilamentos do sistema no Windows.	11
Identificando afunilamentos do sistema no UNIX.	12
Afunilamentos de tempo de execução.	12
Afunilamentos de Otimização de Serviços de Dados SQL.	12
Afunilamentos de Otimização de Serviços da Web.	13
Afunilamentos de Conexões.	13
 Capítulo 2: Otimização de Destinos.....	14
Visão Geral da Otimização de Destinos.	14
Otimização de Destinos de Arquivo Simples.	14
Intervalos de Pontos de Verificação do Banco de Dados.	15
Cargas em Massa.	15
Otimização de Destinos de Banco de Dados.	16
 Capítulo 3: Otimização de Origens.....	17
Visão Geral da Otimização de Origens.	17
Otimização de Origens de Arquivo Simples.	18
Otimização de Consultas.	18
Filtros Condicionais.	19
Selecionar Distintas.	19
Dicas.	20
Regras e Diretrizes de Dicas.	20
Criando Dicas.	21
Restrições.	21

Configurando Restrições.	22
Otimização de Objetos de Dados Personalizados.	22
Otimização de Origens de Banco de Dados.	23
Capítulo 4: Otimização de Transformações.	24
Otimização de Transformações.	24
Otimização de Transformações de Agregador.	24
Otimização de Expressões.	25
Otimização de Transformações Java.	27
Otimização de Seleção Antecipada com a Transformação Java.	27
Otimização de Envio com a Transformação Java.	29
Otimização de Transformações de Associador.	30
Otimização de Transformações de Pesquisa.	30
Otimização de Transformações de Classificador.	33
Otimização de Transformações SQL.	34
Otimização de Seleção Antecipada com a Transformação SQL.	34
Otimização de Envio com a Transformação SQL.	34
Cache de Transformações.	35
Eliminação de Erros de Transformação.	36
Efeitos Colaterais de Transformações.	36
Otimização de Transformações de Consumidor de Serviço da Web.	37
Otimização de Seleção Antecipada com a Transformação de Consumidor de Serviço da Web.	38
Otimização de Envio com a Transformação de Consumidor de Serviço da Web.	38
Capítulo 5: Otimização de Mapeamentos.	41
Visão Geral da Otimização de Mapeamentos.	41
Métodos de Otimização.	42
Níveis do Otimizador.	42
Otimização de Filtro.	43
Método de Otimização de Projeção Antecipada.	43
Método de Otimização de Predicado.	44
Método de Otimização Baseada em Custos.	45
Método de Otimização Dataship-Join.	45
Método de Otimização de Semi-associação.	46
Método de Otimização de Seleção Antecipada.	48
Método de Otimização de Predicado Global.	48
Método de Otimização de Ajuste de Ramificação.	48
Método de Otimização de Envio.	49
Otimização de Empilhamento.	49
Otimização de Empilhamento Total.	50
Empilhamento de Origem.	51
Regras e Diretrizes para Otimização de Empilhamento.	51
Leitura de Passagem Única.	51

Otimização de Filtros.	52
Otimização de Conversão de Tipos de Dados.	52
Rastreamento de Erros.	53
Capítulo 6: Otimização de Mapeamento Particionado.	54
Visão Geral da Otimização de Mapeamento Particionado.	54
Usar Várias CPUs.	55
Aumentar o Valor Máximo de Paralelismo.	55
Otimizar Arquivos Simples para Particionamento.	56
Otimizar Origens de Arquivo Simples para Particionamento.	56
Otimizar Destinos de Arquivo Simples para Particionamento.	56
Otimizar Bancos de Dados Relacionais para Particionamento.	57
Otimizar o Banco de Dados de Origem para Particionamento.	57
Otimizar o Banco de Dados de Destino para Particionamento.	57
Otimizar Transformações para Particionamento.	58
Capítulo 7: Otimização do Tempo de Execução.	60
Visão Geral da Otimização do Tempo de Execução.	60
Otimização de Serviços de Aplicativo.	60
Otimização do Serviço Analyst.	61
Otimização do Serviço de Integração de Dados.	61
Otimização do Serviço de Repositório do Modelo.	62
Estatísticas de Monitoramento.	63
Alocação de Memória.	64
Armazenamento em Cache do Objeto de Dados.	65
Tipos de Dados para Tabelas de Cache.	66
Otimização do Cache de Objetos de Dados.	68
Otimização do Sistema.	68
Capítulo 8: Otimização de Serviços de Dados SQL.	70
Visão Geral da Otimização de Serviço de Dados SQL.	70
Otimização de Ferramentas de Cliente de Terceiros.	71
Níveis do Otimizador do Serviço de Dados SQL.	71
Configurando o Nível do Otimizador do Serviço de Dados SQL para Visualização de Dados . . .	72
Configurando o Nível do Otimizador para Serviços de Dados SQL Implantados.	73
Plano de Consulta do Serviço de Dados SQL.	73
Exibindo um Plano de Consulta SQL.	74
Propriedades do Serviço de Dados SQL para Solicitações Simultâneas e de Memória.	75
Cache de Conjunto de Resultados para um Serviço de Dados SQL.	76
Propriedades do Cache de Conjunto de Resultados do Serviço de Dados SQL.	77
Habilitando o Armazenamento em Cache do Conjunto de Resultados para um Serviço de Dados SQL.	77
Dados Virtuais Persistentes em Tabelas Temporárias.	78

Implementação da Tabela Temporária.	78
Capítulo 9: Otimização de Serviços da Web.	79
Visão Geral da Otimização de Serviços da Web.	79
Otimizar Solicitações HTTP.	80
Compactação de Mensagens de Serviços da Web.	80
Nível do Otimizador de Serviços da Web.	80
Configurando o Nível do Otimizador do Serviço da Web para Visualização de Dados	81
Configurando o Nível do Otimizador para Serviços da Web Implantados.	81
Propriedades dos Serviços da Web para Solicitações Simultâneas e de Memória	82
Exemplo de Configuração do Serviço de Integração de Dados para Solicitações de Serviços da Web Simultâneas	84
Propriedade de Serviço da Web para Configurar uma Instância do DTM Ativo.	85
Armazenamento em Cache de Conjunto de Resultados do Serviço da Web.	85
Habilitando o Armazenamento em Cache de Conjunto de Resultados para um Serviço da Web.	86
Gerenciamento de Logs de Serviços da Web.	86
Capítulo 10: Otimização de Conexões.	87
Visão Geral da Otimização de Conexões.	87
Pool de Conexões.	87
Propriedades do Pool em Objetos de Conexão.	88
Tamanho dos Pacotes de Rede do Banco de Dados.	88
Índice.	90

Prefácio

O *Guia de Ajuste do Desempenho da Informatica* foi elaborado para administradores e desenvolvedores interessados em melhorar o desempenho. Esse guia pressupõe que você conheça os sistemas operacionais, as redes, as ferramentas cliente, os conceitos de bancos de dados relacionais e os arquivos simples no seu ambiente. Para obter informações sobre o ajuste do desempenho de banco de dados que não são abordadas no guia, consulte a documentação que acompanha seus produtos de banco de dados.

Recursos da Informatica

Rede da Informatica

A Rede da Informatica hospeda o Suporte Global a Clientes da Informatica, a Base de Dados de Conhecimento da Informatica e outros recursos de produtos. Para acessar a Rede da Informatica, visite <https://network.informatica.com>.

Como membro, você pode:

- Acessar todos os seus recursos Informatica em um só lugar.
- Pesquisar a Base de Dados de Conhecimento em busca de recursos de produtos, incluindo documentações, perguntas frequentes e práticas recomendadas.
- Visualizar informações sobre disponibilidade de produtos.
- Revisar seus casos de suporte.
- Encontrar a sua Rede de Grupo de Usuários da Informatica local e colaborar com seus colegas.

Base de Dados de Conhecimento da Informatica

Use a Base de Dados de Conhecimento da Informatica para pesquisar a Rede da Informatica em busca de recursos de produtos, como documentações, artigos de instruções, práticas recomendadas e PAMs.

Para acessar a Base de Dados de Conhecimento, visite <https://kb.informatica.com>. Em caso de dúvidas, comentários ou ideias sobre a Base de Dados de Conhecimento, entre em contato com a equipe da Base de Dados de Conhecimento da Informatica em KB_Feedback@informatica.com.

Documentação da Informatica

Para obter a documentação mais recente do seu produto, navegue pela Base de Dados de Conhecimento da Informatica

em https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx

Em caso de dúvidas, comentários ou ideias sobre esta documentação, entre em contato com a equipe de Documentação da Informatica pelo e-mail infa_documentation@informatica.com.

Matrizes de Disponibilidade de Produto Informatica

As Matrizes de Disponibilidade de Produto (PAMs) indicam as versões dos sistemas operacionais, os bancos de dados e outros tipos de fontes e destinos de dados com os quais uma versão de produto é compatível. Se você for membro da Rede da Informatica, poderá acessar PAMs em

<https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

O Informatica Velocity é uma coleção de dicas e práticas recomendadas desenvolvidas pelos Serviços Profissionais da Informatica. Desenvolvido com base na experiência no mundo real de centenas de projetos de gerenciamento de dados, o Informatica Velocity representa o conhecimento coletivo de nossos consultores, que trabalharam com organizações de todo o mundo para planejar, desenvolver, implantar e manter soluções de gerenciamento de dados bem-sucedidas.

Se você for membro da Rede da Informatica, poderá acessar os recursos do Informatica Velocity em

<http://velocity.informatica.com>.

Se você tiver dúvidas, comentários ou ideias sobre o Informatica Velocity, entre em contato com os Serviços Profissionais da Informatica em ips@informatica.com.

Informatica Marketplace

O Informatica Marketplace é um fórum onde você pode encontrar soluções que aumentam, ampliam ou aprimoram suas implementações da Informatica. Aproveitando qualquer uma das centenas de soluções fornecidas por desenvolvedores e parceiros da Informatica, você pode melhorar sua produtividade e agilizar o tempo de implementação nos seus projetos. Você pode acessar o Informatica Marketplace através do link <https://marketplace.informatica.com>.

Suporte global a clientes Informatica

Você pode entrar em contato com um Centro de Suporte Global por telefone ou via Suporte Online na Rede da Informatica.

Para descobrir o número de telefone local do Suporte Global a Clientes da Informatica, visite o site da Informatica no seguinte link:

<http://www.informatica.com/us/services-and-training/support-services/global-support-centers>.

Se você for membro da Rede da Informatica, poderá usar o Suporte Online em

<http://network.informatica.com>.

CAPÍTULO 1

Visão Geral do Ajuste do Desempenho

Este capítulo inclui os seguintes tópicos:

- [Visão Geral do Ajuste do Desempenho, 9](#)
- [Afunilamentos de Destinos, 10](#)
- [Afunilamentos de Origens, 10](#)
- [Afunilamentos de Mapeamentos, 11](#)
- [Afunilamentos do Sistema de Computador, 11](#)
- [Afunilamentos de tempo de execução, 12](#)
- [Afunilamentos de Otimização de Serviços de Dados SQL, 12](#)
- [Afunilamentos de Otimização de Serviços da Web, 13](#)
- [Afunilamentos de Conexões, 13](#)

Visão Geral do Ajuste do Desempenho

O objetivo do ajuste do desempenho é eliminar afunilamentos de desempenho. Um afunilamento é uma área no mapeamento que é executada com mais frequência e apresenta a menor velocidade de processamento. Um afunilamento reduz o desempenho geral do mapeamento.

Para otimizar um mapeamento, identifique um afunilamento de desempenho, elimine-o e depois identifique o próximo afunilamento de desempenho. Otimize um componente de mapeamento de cada vez. Você pode cronometrar um mapeamento antes e depois da alteração para verificar se a otimização apresenta um impacto no desempenho.

Você pode otimizar os componentes de mapeamento na seguinte ordem:

1. Destinos
2. Origens
3. Mapeamentos
4. Transformações
5. Ambiente Informatica na ferramenta Administrator
6. O sistema de computador
7. Serviço de Dados ou serviço da Web

Use os seguintes métodos para identificar os afunilamentos de desempenho:

- Executar mapeamentos de teste. Você pode configurar um mapeamento de teste para ler dados de uma origem de arquivo simples ou gravar dados em um destino de arquivo simples para identificar afunilamentos de origens e destinos.
- Analisar detalhes de desempenho. Analise detalhes de desempenho, como métodos de otimização, para determinar onde o desempenho do mapeamento diminui.
- Monitorar o desempenho do sistema. Você pode usar ferramentas de monitoramento do sistema para exibir a porcentagem de uso da CPU, esperas de E/S, a paginação e o uso de recursos do sistema.

Afunilamentos de Destinos

Afunilamentos de destinos são quedas de desempenho quando o Data Integration Service grava em um destino. Afunilamentos de destinos podem ocorrer quando o banco de dados usa intervalos curtos de pontos de verificação ou tamanhos pequenos de pacotes de rede de banco de dados.

O afunilamento de desempenho mais comum ocorre quando o Data Integration Service grava em um banco de dados de destino. Se o banco de dados usar intervalos curtos de pontos de verificação, seu processamento ficará lento com mais frequência quando ocorrer uma gravação em um ponto de verificação. Tamanhos pequenos de pacotes de rede de banco de dados podem causar afunilamentos. É possível permitir que pacotes maiores de dados atravessem a rede de uma só vez.

Para identificar um afunilamento de destino, você pode criar uma cópia do mapeamento que possui um destino de arquivo simples em vez de um destino de banco de dados. Se o desempenho aumentar significativamente, significa que existe um afunilamento de destino. Se o mapeamento já grava dados em um destino de arquivo simples, você provavelmente não tem um afunilamento de destino.

Afunilamentos de Origens

Afunilamentos de origens são quedas de desempenho quando o Data Integration Service lê dados de um banco de dados de origem. Afunilamentos de origens podem ocorrer quando a consulta de origem não é eficiente ou quando os tamanhos dos pacotes de rede de banco de dados são pequenos.

Quando o mapeamento lê dados de uma origem relacional, você pode usar os seguintes métodos para identificar afunilamentos de origens:

- Adicionar uma transformação de Filtro ao mapeamento. Adicione a transformação de Filtro depois da origem. Defina a condição de Filtro como falsa para que a transformação de Filtro não retorne dados. Se o tempo necessário pelo mapeamento for o mesmo, então o mapeamento tem um afunilamento de origem.
- Criar um mapeamento de teste de leitura. Faça uma cópia do mapeamento, mas remova todas as transformações, associações ou consultas. Conecte a origem a um destino. Se o desempenho do desempenho for semelhante ao mapeamento original, significa que existe um afunilamento de origem.
- Executar a consulta de leitura diretamente no banco de dados de origem. Copie a consulta de leitura a partir do log de mapeamento. Execute a consulta no banco de dados de origem com uma ferramenta de consulta, como o isql. Meça o tempo de execução e o tempo necessário para a consulta retornar uma linha.

Afunilamentos de Mapeamentos

Se você determinar que não existe um afunilamento de origem ou de destino, talvez exista um afunilamento de mapeamento. Um tamanho pequeno de cache, uma memória de buffer insuficiente e intervalos de confirmação curtos podem causar afunilamentos de mapeamento.

Para identificar um afunilamento de mapeamento, analise os detalhes do desempenho no log de mapeamento. Os detalhes do desempenho incluem informações sobre cada transformação, como o número de linhas de entrada, linhas de saída e linhas de erro.

Você também pode adicionar uma transformação de Filtro antes de cada definição de destino. Defina a condição de Filtro como falsa para que a transformação de Filtro não carregue dados nas tabelas de destino. Se o tempo necessário para executar o novo mapeamento for igual ao mapeamento original, significa que existe um afunilamento de mapeamento.

Afunilamentos do Sistema de Computador

Você pode exibir o uso de recursos ao executar serviços Informatica no Windows ou no UNIX. No Windows, use o Gerenciador de Tarefas. O UNIX tem várias ferramentas que você pode usar para analisar o desempenho.

Identificando afunilamentos do sistema no Windows

Você pode visualizar a guia Desempenho e processos no Gerenciador de Tarefas para obter informações do sistema. A guia Desempenho no Gerenciador de Tarefas oferece uma visão geral do uso da CPU e do total de memória usada. Use o Monitor de Desempenho para visualizar mais informações detalhadas.

A tabela a seguir descreve as informações de sistema que você pode usar no Monitor de Desempenho do Windows para criar um gráfico:

Propriedade	Descrição
Tempo do processador em percentual	Se você tiver mais de uma CPU, monitore cada uma delas para tempo do processador em percentual.
Páginas/segundo	Se as páginas/segundo for maior do que cinco, talvez você tenha pressão de memória excessiva conhecido como sobrecarga.
Tempo em percentual dos discos físicos	O percentual de tempo que o disco físico está ocupado executando solicitações de leitura e gravação.
Comprimento da fila dos discos físicos	O número de usuários aguardando para acessar o mesmo dispositivo de disco.
Total de bytes do servidor por segundo	O servidor enviou e recebeu da rede.

Identificando afunilamentos do sistema no UNIX

Use as seguintes ferramentas para identificar afunilamentos do sistema no UNIX:

- **superior.** Visualize o desempenho geral do sistema. Esta ferramenta exibe o uso da CPU, da memória e de permuta do sistema e de processos individuais em execução no sistema.
- **iostat.** Monitore a operação de carregamento para cada disco anexado ao servidor do banco de dados. O iostat exibe a porcentagem de tempo que o disco está fisicamente ativo. Se você utiliza matrizes de disco, use os utilitários fornecidos com as matrizes em vez do iostat.
- **vmstat.** Monitore as ações de permuta do disco.
- **sar.** Visualize relatórios detalhados das atividades de sistema da CPU, da memória e do uso do disco. Você pode usar esta ferramenta para monitorar o carregamento da CPU. Ela oferece o uso em percentual com base no tempo de espera, no tempo ocioso, no sistema e no usuário. Além disso, use esta ferramenta para monitorar as ações de permuta do disco.

Afunilamentos de tempo de execução

Ative recursos de desempenho e ajuste as propriedades do Serviço de Integração de Dados para otimizar o desempenho de mapeamentos. Defina configurações de otimização para o Serviço de Integração de Dados e o Serviço de Repositório do Modelo na ferramenta Administrator.

Aloque memória para obter o desempenho ideal do sistema e configure níveis de rastreamento de erros para reduzir o número de eventos de log gerados pelo Serviço de Integração de Dados quando ele executa o mapeamento.

Você pode configurar a quantidade máxima de memória alocada pelo Serviço de Integração de Dados para a execução de todas as solicitações simultâneas. Também é possível limitar a quantidade máxima de memória alocada pelo Serviço de Integração de Dados para qualquer solicitação especificada.

Você pode configurar o cache do conjunto de resultados para permitir que o Serviço de Integração de Dados armazene em cache os resultados do processo do DTM associado a cada consulta de serviço de dados SQL e solicitação de serviço da Web.

Afunilamentos de Otimização de Serviços de Dados SQL

Você pode otimizar serviços de dados SQL para melhorar o desempenho quando os usuários finais executam consultas SQL nesses serviços usando ferramentas de cliente de terceiros. Se um serviço de dados SQL usar um mapeamento de tabela virtual, você poderá otimizar as transformações e o mapeamento.

É possível otimizar o driver JDBC para melhorar o desempenho durante a consulta de um serviço de dados SQL. Também é possível configurar o cache de objetos de dados para o Data Integration Service melhorar o desempenho de mapeamentos e consultas SQL.

Afunilamentos de Otimização de Serviços da Web

Você pode otimizar serviços da Web para melhorar o desempenho quando o Data Integration Service executa solicitações de serviços da Web. Ajuste o Data Integration Service para gerenciar a memória, manipular solicitações de serviços da Web simultâneas e manter um processo do DTM ativo para que ele possa processar mais de uma solicitação de serviço da Web.

Para melhorar o desempenho de serviços da Web, use a compactação de mensagens de serviços da Web, otimize solicitações HTTP e configure o cache de objetos de dados.

Afunilamentos de Conexões

Você pode otimizar conexões para melhorar o desempenho. É possível gerenciar o pool de instâncias de conexão inativas para uma conexão de banco de dados. Além disso, você pode aumentar o tamanho dos pacotes de rede para permitir que pacotes de dados maiores atravessem a rede de uma vez.

CAPÍTULO 2

Otimização de Destinos

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Otimização de Destinos, 14](#)
- [Otimização de Destinos de Arquivo Simples, 14](#)
- [Intervalos de Pontos de Verificação do Banco de Dados, 15](#)
- [Cargas em Massa, 15](#)
- [Otimização de Destinos de Banco de Dados, 16](#)

Visão Geral da Otimização de Destinos

Otimize destinos para permitir que o Serviço de Integração de Dados grave nesses destinos eficientemente. Você pode descartar índices e restrições de chave antes de executar um mapeamento, aumentar o número de intervalos de ponto de verificação no banco de dados, configurar o carregamento em massa nas propriedades de gravação para um objeto de dados e otimizar um banco de dados de destino Oracle.

Use as seguintes técnicas de otimização para otimizar o destino:

- Otimizar destinos de arquivo simples.
- Aumentar intervalos de ponto de verificação do banco de dados.
- Usar cargas em massa.
- Otimizar bancos de dados de destino Oracle.

Otimização de Destinos de Arquivo Simples

Você pode melhorar o desempenho de mapeamento otimizando destinos de arquivo simples. Também pode enviar tarefas de transformação a um comando para melhorar o desempenho.

Considere as seguintes soluções para reduzir afunilamentos de destinos de arquivo simples:

Enviar tarefas de transformação para um comando, e não para o Data Integration Service.

Você pode melhorar o desempenho de mapeamento enviando tarefas de transformação para um comando, e não para o Data Integration Service. Também pode usar um comando para classificar ou compactar dados de destino. Na ferramenta Developer, configure a propriedade Command nas propriedades de tempo de execução para um destino de arquivo simples.

No UNIX, use qualquer script de shell ou comando UNIX válido. No Windows, use qualquer arquivo em lote ou de comando do DOS válido. O gravador de arquivo simples envia os dados ao comando em vez de enviá-los a um destino de arquivo simples.

Por exemplo, use o comando a seguir para gerar um arquivo compactado a partir dos dados de destino:

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

Gravar em um destino de arquivo simples local para o nó de processo do serviço.

Se o Data Integration Service for executado em um único nó e gravar em um destino de arquivo simples, você poderá otimizar o desempenho de mapeamento gravando em um destino de arquivo simples local para o nó de processo do serviço.

Intervalos de Pontos de Verificação do Banco de Dados

O desempenho do Data Integration Service diminui sempre que ele precisa aguardar enquanto o banco de dados realiza um ponto de verificação.

Considere a seguinte solução para reduzir afunilamentos de ponto de verificação do banco de dados:

Aumentar o intervalo de pontos de verificação no banco de dados.

Para diminuir o número de pontos de verificação e melhorar o desempenho, aumente o intervalo de pontos de verificação no banco de dados.

Embora você aumente desempenho ao reduzir o número de pontos de verificação, também aumentará o tempo de recuperação se o banco de dados for desligado inesperadamente.

Cargas em Massa

Quando você usa o carregamento em massa, o Data Integration Service ignora o log do banco de dados, o que acelera o desempenho.

Considere as seguintes soluções para reduzir afunilamentos de carregamento em massa:

Configurar o carregamento em massa nas propriedades de gravação de um objeto de dados.

É possível usar o carregamento em massa para melhorar o desempenho de um mapeamento que insere uma grande quantidade de dados em um banco de dados DB2, Sybase ASE, Oracle ou Microsoft SQL Server.

Entretanto, sem gravar no log do banco de dados, o banco de dados de destino não pode realizar a reversão. Como resultado, talvez não seja possível realizar uma recuperação. Quando for usar o carregamento em massa, avalie a importância de um maior desempenho de mapeamento versus a capacidade de recuperar um mapeamento incompleto.

Otimização de Destinos de Banco de Dados

Você pode otimizar o banco de dados de destino verificando a cláusula de armazenamento, a alocação de espaço e segmentos de reversão ou cancelamento de ação.

Considere as seguintes soluções para reduzir afunilamentos de destinos de banco de dados:

Verifique se o banco de dados armazena segmentos de reversão ou cancelamento de ação em espaços de tabela adequados, preferencialmente em discos diferentes.

Quando você grava no banco de dados, este usa segmentos de reversão ou cancelamento de ação durante os carregamentos. Consulte o administrador do banco de dados para verificar se o banco de dados armazena segmentos de reversão ou cancelamento de ação em espaços de tabela adequados, preferencialmente em discos diferentes. Os segmentos de reversão ou cancelamento de ação devem também ter cláusulas de armazenamento apropriadas.

Ajustar o log de repetição de ação do banco de dados.

Para otimizar o banco de dados, ajuste o log de repetição de ação. O banco de dados usa o log de repetição de ação para registrar operações de carregamento. Certifique-se de que o tamanho do log de repetição de ação e o tamanho do buffer sejam ideais. Para um banco de dados Oracle, você pode exibir as propriedades do log de repetição de ação no arquivo init.ora.

Conectar-se a um banco de dados Oracle com o protocolo IPC.

Se o Data Integration Service for executado em um único nó, e a instância Oracle for local para o nó de processo do serviço, será possível otimizar o desempenho usando o protocolo IPC para conexão com o banco de dados Oracle. Você pode configurar a conexão de banco de dados Oracle em listener.ora e em tnsnames.ora.

CAPÍTULO 3

Otimização de Origens

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Otimização de Origens, 17](#)
- [Otimização de Origens de Arquivo Simples, 18](#)
- [Otimização de Consultas, 18](#)
- [Filtros Condicionais, 19](#)
- [Selecionar Distintas, 19](#)
- [Dicas, 20](#)
- [Restrições, 21](#)
- [Otimização de Objetos de Dados Personalizados, 22](#)
- [Otimização de Origens de Banco de Dados, 23](#)

Visão Geral da Otimização de Origens

Otimize fontes de dados de arquivo simples, relacionais e personalizadas para permitir que o Serviço de Integração de Dados leia dados de origem de maneira eficiente.

Use as seguintes técnicas de otimização para otimizar origens:

- Ler dados de origem eficientemente.
- Usar técnicas de otimização de consultas.
- Use filtros condicionais com a consulta SQL.
- Selecionar valores exclusivos da origem.
- Aplicar dicas à consulta SQL.
- Configure restrições em objetos de dados lógicos, objetos de dados físicos e tabelas virtuais.
- Configurar objetos de dados personalizados para otimização.
- Configurar bancos de dados Oracle, Sybase e Microsoft SQL Server para otimização.

Otimização de Origens de Arquivo Simples

Configure as propriedades de formato para origens de arquivo simples para permitir que o Data Integration Service leia dados de origem de maneira eficiente.

Considere as seguintes soluções para afunilamentos de origens de arquivo simples:

Não usar aspas ou caracteres de escape nas propriedades de formato para um arquivo simples delimitado.

Se você especificar um caractere de escape, o Data Integration Service lerá o caractere delimitador como um caractere comum inserido na string. É possível melhorar um pouco o desempenho do mapeamento quando o arquivo de origem não contém aspas ou caracteres de escape.

Definir o número de bytes que o Data Integration Service lê por linha.

Se o mapeamento for lido a partir de uma origem de arquivo simples, você poderá melhorar o desempenho do mapeamento definindo o número de bytes lidos por linha pelo Data Integration Service. Configure a propriedade Comprimento do Buffer Sequencial em Linha nas propriedades de tempo de execução para origens de arquivo simples.

Por padrão, o Data Integration Service lê 1024 bytes por linha. Se cada linha no arquivo de origem for menor que a configuração padrão, será possível reduzir o comprimento do buffer sequencial em linha nas propriedades do mapeamento.

Otimização de Consultas

Se um mapeamento associar várias tabelas de origem em um objeto de dados personalizado, você poderá melhorar o desempenho otimizando a consulta com dicas de otimização. Além disso, instruções únicas de seleção de tabela com uma cláusula ORDER BY ou GROUP BY podem se beneficiar com a otimização, como a adição de índices.

Considere as seguintes soluções para afunilamentos de consultas:

Criar dicas do otimizador para informar ao banco de dados como executar a consulta para um determinado conjunto de tabelas de origem.

Em geral, o otimizador de banco de dados determina a maneira mais eficiente de processar os dados da origem. No entanto, talvez conheça propriedades sobre as tabelas de origem que o otimizador de banco de dados não conhece. O administrador de banco de dados pode criar dicas do otimizador para informar ao banco de dados como executar a consulta para um determinado conjunto de tabelas de origem.

Configurar dicas do otimizador para começar a retornar linhas o mais rápido possível, ao invés de retornar todas as linhas de uma vez.

Use dicas de otimização se houver um longo atraso entre o início da execução da consulta e o momento em que o Data Integration Service recebe a primeira linha de dados. Configurar dicas do otimizador para começar a retornar linhas o mais rápido possível, ao invés de retornar todas as linhas de uma vez. Isso permite que o Data Integration Service processe as linhas em paralelo com a execução da consulta.

Criar um índice nas colunas ORDER BY ou GROUP BY.

Consultas que contêm cláusulas ORDER BY ou GROUP BY podem se beneficiar com a criação de um índice nas colunas ORDER BY ou GROUP BY. Depois de otimizar a consulta, use a opção de substituição SQL para tirar proveito total dessas modificações.

Configurar o banco de dados para executar consultas paralelas.

Você também pode configurar o banco de dados de origem de forma a executar consultas paralelas para melhorar o desempenho. Para obter mais informações sobre como configurar consultas paralelas, consulte a documentação do banco de dados.

A consulta que o Data Integration Service utiliza para ler dados aparece no banco de dados virtual em um Serviço de Dados SQL. Você também pode localizar a consulta no objeto de dados personalizado. Faça com que o administrador do banco de dados analise a consulta e depois crie índices e dicas do otimizador para as tabelas de origem.

Filtros Condicionais

Às vezes, um filtro de origem simples no banco de dados de origem pode influenciar de maneira negativa o desempenho devido à falta de índices. Você pode usar o filtro condicional no objeto de dados personalizado para melhorar o desempenho.

Considere a seguinte solução para afunilamentos de filtros condicionais:

Usar o filtro condicional para vários mapeamentos que lêem a partir da mesma origem simultaneamente.

Se vários mapeamentos fizerem leituras a partir da mesma origem simultaneamente, o filtro condicional poderá melhorar o desempenho.

No entanto, alguns mapeamentos poderão apresentar um desempenho mais rápido se os dados de origem forem filtrados no banco de dados de origem. Você pode testar o mapeamento com o filtro de banco de dados e com o filtro condicional para determinar qual método melhora o desempenho.

Selecionar Distintas

É possível selecionar valores exclusivos de origens em um objeto de dados personalizado usando a opção Selecionar Distintas. Quando você usa Selecionar Distintas, o Data Integration Service adiciona uma instrução `SELECT DISTINCT` à consulta SQL padrão.

Considere a seguinte solução para afunilamentos em Selecionar Distintas:

Usar a opção Selecionar Distintas para filtrar dados desnecessários com antecedência no fluxo de dados.

Use a opção Selecionar Distintas para o objeto de dados personalizado se quiser que o Data Integration Service selecione valores exclusivos de uma origem. Usar a opção Selecionar Distintas para filtrar dados desnecessários com antecedência no fluxo de dados. Isso pode melhorar o desempenho.

Por exemplo, você pode usar a opção Selecionar Distintas para extrair IDs de cliente exclusivos de uma tabela que lista o total de vendas. Quando você usa o objeto de dados personalizado em um mapeamento, o Data Integration Service remove dados desnecessários com maior antecedência no fluxo de dados, o que pode melhorar o desempenho.

Dicas

Você pode adicionar dicas à consulta SQL de origem para passar instruções para um otimizador de banco de dados. O otimizador usa as dicas para escolher um plano de execução de consulta para acessar a origem.

O campo Dicas aparece na exibição **Consulta** de uma instância de objeto de dados relacionais ou de um objeto de dados personalizados. O banco de dados de origem deve ser Oracle, Sybase, IBM DB2 ou Microsoft SQL Server. O campo Dicas não aparece para outros tipos de banco de dados.

Ao gerar a consulta de origem, o Data Integration Service adiciona as dicas SQL à consulta exatamente como você a insere na ferramenta Developer. O Data Integration Service não analisa as dicas. Quando você executa o mapeamento que contém a origem, o log de mapeamento mostra a consulta com as dicas na consulta.

O Data Integration Service insere as dicas SQL em uma posição na consulta dependendo do tipo de banco de dados. Consulte a documentação do banco de dados para obter informações sobre a sintaxe de dicas.

Oracle

O Data Integration Service adiciona dicas diretamente depois da palavra-chave SELECT/UPDATE/INSERT/DELETE.

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

'+' indica o início das dicas.

As dicas estão contidas em um comentário (/*... */ ou --... até o final da linha)

Sybase

O Data Integration Service adiciona dicas após a consulta. Configure um nome de plano na dica.

```
SELECT ... PLAN <plano>
```

```
selecionar avg(price) no plano de títulos "(scalar_agg (i_scan type_price_ix titles )"
```

IBM DB2

É possível a cláusula optimize-for como dica. O Data Integration Service adiciona a cláusula no final da consulta.

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

A cláusula optimize-for informa ao otimizador de banco de dados quantas linhas a consulta pode processar. A cláusula não limita o número de linhas. Se o banco de dados processar mais de <n> linhas, o desempenho poderá diminuir.

Microsoft SQL Server

O Data Integration Service adiciona dicas ao final da consulta como parte de uma cláusula OPTION.

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

Regras e Diretrizes de Dicas

Use as seguintes regras e diretrizes ao configurar dicas para consultas SQL:

- Se você habilitar a otimização de empilhamento ou se usar uma semi-associação em um objeto de dados relacionais, a consulta de origem original será alterada. O Data Integration Service não aplica dicas à consulta modificada.

- Você pode combinar dicas com substituições de associação e filtro, mas, se você configurar uma substituição SQL, esta terá precedência, e o Data Integration Service não aplicará as outras substituições.
- A exibição **Consulta** mostra uma exibição simples ou avançada. Se você inserir uma dica com uma substituição de filtro, classificação ou associação na exibição simples, a ferramenta Developer mostrará a substituição de consulta completa na exibição avançada.

Criando Dicas

Crie dicas para enviar instruções ao otimizador de banco de dados para determinar um plano de consulta.

1. Abra o objeto de dados personalizados ou a instância do objeto de dados relacionais.
2. Selecione a exibição **Leitura**.
3. Selecione a transformação de Saída.
4. Selecione as propriedades de **Consulta**.
5. Selecione a consulta simples.
6. Clique em **Editar** ao lado do campo **Dicas**.
A caixa de diálogo **Dicas** é exibida.
7. Insira a dica no campo **Consulta SQL**.
A ferramenta Developer não valida a dica.
8. Clique em **OK**.
9. Salve o objeto de dados.

Restrições

O Serviço de Integração de Dados pode ler restrições de origens relacionais, origens de arquivo simples, objetos de dados lógicos ou tabelas virtuais. Uma restrição é uma expressão condicional que os valores em uma linha de dados devem satisfazer.

Quando o Serviço de Integração de Dados lê restrições, ele pode descartar as linhas que não são avaliadas como TRUE para as linhas de dados com base no método de otimização aplicado.

Antes de definir uma restrição, verifique se os dados de origem satisfazem a condição definida pela restrição.

Por exemplo, um banco de dados de origem tem uma coluna AGE que é exibida para ter linhas com AGE < 70. Você pode definir uma restrição com AGE < 70 no banco de dados de origem. A Integração de Dados lê os registros do banco de dados de origem com a restrição AGE < 70. Se o Serviço de Integração de Dados ler os registros com AGE > = 70, ele poderá descartar as linhas com AGE > = 70.

No banco de dados, você pode usar comandos SQL para definir restrições no ambiente de banco de dados quando você se conectar ao banco de dados. O Serviço de Integração de Dados executa o SQL de ambiente de conexão sempre que se conecta ao banco de dados.

Use a ferramenta Developer para definir restrições em objetos de dados lógicos, objetos de dados físicos e tabelas virtuais. Quando você definir uma restrição, insira uma expressão avaliada como TRUE para cada linha de dados.

Configurando Restrições

Você pode adicionar restrições a objetos de dados relacionais, objetos de dados de arquivo simples, objetos de dados personalizados, objeto de dados lógicos e tabelas virtuais. Depois que você adicionar uma restrição, poderá editá-la ou excluí-la.

1. Na exibição **Object Explorer**, abra o mapeamento que contém o objeto de dados relacionais adicionado como uma transformação de Leitura. Ou, abra o objeto de dados de arquivo simples, o objeto de dados personalizados, o objeto de dados lógicos ou a tabela virtual.
 - Para definir restrições em um objeto de dados relacionais adicionado a um mapeamento como uma transformação de Leitura, selecione a transformação de Leitura nesse mapeamento. Na exibição **Propriedades**, selecione a guia **Avançado**.
 - Para definir restrições em um objeto de dados de arquivo simples, selecione a exibição **Avançado** e expanda a seção **Tempo de execução: Leitura**.
 - Para definir restrições em um objeto de dados personalizados, selecione a exibição **Leitura** e selecione a porta **Saída** da transformação de origem. Na exibição **Propriedades**, selecione a guia **Avançado**.
 - Para definir restrições em um objeto de dados lógicos, selecione um modelo de dados lógicos e selecione o objeto de dados lógicos. Na exibição **Propriedades**, selecione a guia **Avançado**.
 - Para definir restrições em uma tabela virtual, abra-a no ponto de extremidade SQL. Na exibição **Propriedades**, selecione a guia **Avançado**.
2. Clique no campo de valor das restrições.
A caixa de diálogo **Restrições** é exibida.
3. Clique em **Novo** para abrir o Editor de Expressão.
4. Configure a restrição lógica e use as funções de expressão e as colunas como parâmetros.
5. Clique em **Validar**.
6. Clique em **OK**.

Otimização de Objetos de Dados Personalizados

Você pode configurar objetos de dados personalizados para melhorar o desempenho. Você pode otimizar a consulta SQL, usar filtros condicionais e selecionar valores distintos a partir da origem em um objeto de dados personalizado.

Considere as seguintes soluções para afunilamentos de objetos de dados personalizados:

Criar uma consulta personalizada de forma a emitir uma instrução SELECT especial para o Serviço de Integração de Dados ler dados de origem.

A consulta personalizada substitui a consulta padrão usada pelo Serviço de Integração de Dados para ler dados de origens.

Filtrar linhas quando o Serviço de Integração de Dados ler dados de origem.

Se você incluir uma condição de filtro, o Serviço de Integração de Dados adicionará uma cláusula WHERE à consulta padrão.

Selecionar valores distintos da origem.

Se você usar Selecionar Distintas, o Serviço de Integração de Dados adicionará uma instrução SELECT DISTINCT à consulta SQL padrão.

Aplicar dicas de banco de dados.

Você pode adicionar dicas à consulta SQL de origem para passar instruções para um otimizador de banco de dados.

Configure restrições nos dados de origem.

Se você configurar restrições em arquivos simples e tabelas relacionais em um objeto de dados personalizados, o Serviço de Integração de Dados descartará as linhas não avaliadas como TRUE para as linhas de dados.

Otimização de Origens de Banco de Dados

Se o banco de dados de origem for Oracle, você poderá otimizar o desempenho do Data Integration Service usando o protocolo IPC para se conectar ao banco de dados Oracle. Você também pode mover o banco de dados temporário para uma matriz de disco para melhorar o desempenho.

Considere as seguintes soluções para afunilamentos de origens de banco de dados:

Usar o protocolo IPC para se conectar ao banco de dados Oracle.

Se o Data Integration Service for executado em um único nó, e a instância Oracle for local para o nó de processo do serviço, será possível otimizar o desempenho usando o protocolo IPC para conexão com o banco de dados Oracle. Você pode configurar a conexão de banco de dados Oracle em listener.ora e em tnsnames.ora.

Mover o banco de dados temporário e logs de repetição de ação para uma matriz de disco ou unidades mais velozes.

Ao associar tabelas grandes em um banco de dados, você pode usar uma RAID (matriz redundante de discos independentes) para a localização do cache. Como alternativa, é possível adicionar mais arquivos ao grupo de arquivos primário em outros discos para dividir a carga entre esses discos.

CAPÍTULO 4

Otimização de Transformações

Este capítulo inclui os seguintes tópicos:

- [Otimização de Transformações, 24](#)
- [Otimização de Transformações de Agregador, 24](#)
- [Otimização de Expressões, 25](#)
- [Otimização de Transformações Java, 27](#)
- [Otimização de Transformações de Associador, 30](#)
- [Otimização de Transformações de Pesquisa, 30](#)
- [Otimização de Transformações de Classificador, 33](#)
- [Otimização de Transformações SQL, 34](#)
- [Cache de Transformações, 35](#)
- [Eliminação de Erros de Transformação, 36](#)
- [Efeitos Colaterais de Transformações, 36](#)
- [Otimização de Transformações de Consumidor de Serviço da Web, 37](#)

Otimização de Transformações

Otimize transformações para permitir que o Data Integration Service processe transformações em um mapeamento de maneira eficiente.

Use as seguintes técnicas de otimização para otimizar a transformação:

- Configurar transformações para otimização.
- Eliminar erros de transformação.
- Configurar o cache de transformações.

Otimização de Transformações de Agregador

Transformações de Agregador costumam reduzir o desempenho porque precisam agrupar dados antes de processá-los. Transformações de Agregador precisam de memória adicional para manter resultados de grupos intermediários.

Considere as seguintes soluções para afunilamentos de transformações de Agregador:

Agrupar por colunas simples.

Você pode otimizar transformações de Agregador ao agrupar por colunas simples. Quando possível, use números em vez de strings e datas nas colunas usadas para GROUP BY. Evite expressões complexas nas expressões de Agregador.

Usar uma entrada classificada.

Para melhorar o desempenho do mapeamento, classifique dados para a transformação de Agregador. Use a opção Entrada Classificada para classificar dados.

A opção Entrada Classificada diminui o uso de caches agregados. Quando você usa a opção Entrada Classificada, o Serviço de Integração de Dados considera que todos os dados são classificados por grupo. À medida que o Serviço de Integração de Dados lê linhas para um grupo, ele realiza cálculos de agregação. Quando necessário, ele armazena as informações do grupo na memória.

A opção Entrada Classificada reduz a quantidade de dados armazenados no cache durante o mapeamento e melhora o desempenho. Use a opção Entrada Classificada ou uma transformação de Classificador para transmitir dados classificados para a transformação de Agregador.

Você pode melhorar o desempenho ao usar a opção Entrada Classificada em mapeamentos com várias partições.

Filtrar dados antes de agregá-los.

Se você utilizar uma transformação de Filtro no mapeamento, insira-a antes da transformação de Agregador para reduzir uma agregação desnecessária.

Limitar conexões de portas.

Limite o número de portas de entrada/saída ou de saída conectadas para reduzir a quantidade de dados que a transformação de Agregador armazena no cache de dados.

Otimização de Expressões

Algumas expressões usadas em uma transformação podem diminuir o desempenho.

Considere as seguintes soluções para afinamentos de expressões:

Isolar expressões lentas.

Expressões lentas diminuem o desempenho do mapeamento. Para isolar expressões lentas, remova uma expressão de cada vez do mapeamento e execute o mapeamento para determinar o tempo necessário para essa execução sem a expressão. Se houver uma diferença significativa no tempo de execução do mapeamento, procure maneiras de otimizar a expressão lenta.

Conclua as etapas a seguir para avaliar o desempenho da expressão:

1. Cronografe o mapeamento com as expressões originais.
2. Copie o mapeamento e substitua metade das expressões complexas por uma constante.
3. Execute e cronografe o mapeamento editado.
4. Faça outra cópia do mapeamento e substitua a outra metade das expressões complexas por uma constante.
5. Execute e cronografe o mapeamento editado.

Fatorar a lógica comum.

Se o mapeamento realizar a mesma tarefa em vários locais, reduza o número de vezes que ele realiza essa tarefa movendo-a para um ponto anterior no mapeamento. Por exemplo, você tem um

mapeamento com cinco tabelas de destino. Cada destino requer uma pesquisa de número de CPF. Em vez de realizar a pesquisa cinco vezes, insira a transformação de Pesquisa no mapeamento antes da divisão do fluxo de dados. Em seguida, transmita os resultados da pesquisa para todos os cinco destinos.

Minimizar chamadas de função de agregação.

Quando você escrever expressões, fatore o máximo possível de chamadas de função de agregação. Sempre que você usa uma chamada de função de agregação, o Data Integration Service precisa pesquisar e agrupar os dados. Por exemplo, na expressão a seguir, o Data Integration Service lê COLUMN_A, localiza a soma, lê COLUMN_B, localiza a soma e, por fim, localiza a soma das duas colunas:

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

Se você fatorar a chamada de função de agregação, conforme abaixo, o Data Integration Service adicionará COLUMN_A a COLUMN_B e, em seguida, localizará a soma de ambas.

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

Substituir expressões comuns por variáveis locais.

Se você usar a mesma expressão várias vezes em uma transformação, será possível transformar essa expressão em uma variável local. Você pode usar variáveis locais somente dentro da transformação. No entanto, se você calcular a variável somente uma vez, o desempenho melhorará.

Escolher operadores numéricos versus operadores de string.

O Data Integration Service processa operações numéricas com mais rapidez do que operações de string. Por exemplo, se você pesquisar grandes quantidades de dados em duas colunas, EMPLOYEE_NAME e EMPLOYEE_ID, a configuração da pesquisa em EMPLOYEE_ID melhorará o desempenho.

Otimizar comparações CHAR-CHAR e CHAR-VARCHAR.

Quando o Data Integration Service realiza comparações entre as colunas CHAR e VARCHAR, ele fica mais lento toda vez que encontra espaços em branco à direita/esquerda na linha. Você pode usar a opção TreatCHARasCHARonRead ao configurar o Data Integration Service no Informatica Administrator para que o Data Integration Service não ajuste os espaços em branco à direita na extremidade dos campos de origem Char.

Escolha DECODE versus LOOKUP.

Quando você usa a função LOOKUP, o Data Integration Service precisa pesquisar uma tabela em um banco de dados. Ao usar uma função DECODE, você incorpora os valores de pesquisa à expressão para que o Data Integration Service não precise pesquisar uma tabela separada. Sendo assim, quando quiser pesquisar um pequeno conjunto de valores imutáveis, use DECODE para melhorar o desempenho.

Usar operadores em vez de funções.

O Data Integration Service lê expressões gravadas com operadores com mais rapidez do que expressões com funções. Quando possível, use operadores para gravar expressões. Por exemplo, você tem a seguinte expressão que contém funções CONCAT aninhadas:

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

Você pode regravar essa expressão com o operador || da seguinte maneira:

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

Otimizar funções IIF.

Funções IIF podem retornar um valor e uma ação, o que permite expressões mais compactas. Por exemplo, você tem uma origem com três sinalizadores S/N: FLG_A, FLG_B, FLG_C. Você deseja retornar os valores com base nos valores de cada sinalizador.

Você usa a seguinte expressão:

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

Otimização de Transformações Java

Algumas transformações Java em um mapeamento podem diminuir o desempenho.

Considere a seguinte solução para melhorar o desempenho de transformações Java:

Habilitar métodos de otimização com seleção antecipada ou de otimização de filtros de envio com a transformação Java.

Você pode habilitar a otimização com seleção antecipada ou a otimização de envio em transformações Java. Atualize os trechos de código na guia **Interfaces do Otimizador** da transformação Java.

Otimização de Seleção Antecipada com a Transformação Java

Você pode habilitar uma transformação Java ativa ou passiva para otimização com seleção antecipada quando essa transformação não tem efeitos colaterais. O otimizador transmite a lógica de filtro através da transformação Java e modifica a condição de filtro conforme necessário.

Para exibir os trechos de código para otimização com seleção antecipada, escolha PredicatePushOptimization no navegador da guia **Interfaces do Otimizador**.

allowPredicatePush

Booleano. Permite a seleção antecipada. Altere a função para retornar um resultado "true" e a mensagem para habilitar a seleção antecipada. O padrão é "false", e a função retorna uma mensagem informando que não há suporte para a otimização.

```
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 se a transformação Java pode retornar um erro de campo de saída, como um erro de divisão por zero. Altere a função para retornar "false" se a transformação Java não gerar erros de campos de

saída. Quando a transformação Java pode gerar erros de campo, o Data Integration Service não pode usar a otimização com seleção antecipada.

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

getInputExpr

Retorna uma expressão da Informatica que descreve quais valores de entrada de campos de entrada formam um campo de saída. O otimizador precisa saber quais campos de entrada formam um campo de saída para enviar a lógica de filtro pela transformação.

```
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 exemplo, um mapeamento contém uma expressão de filtro, "out0 > 8". Out0 é o valor da porta de saída out0 na transformação Java. Você pode definir o valor de out0 como o valor da porta de entrada in0 + 5. O otimizador pode enviar a seguinte expressão "(in0 + 5) > 8" após a transformação Java com otimização de seleção antecipada. Você poderá retornar NULL se um campo de saída não tiver uma expressão de campo. O otimizador não envia expressões de filtro após campos de saída sem uma expressão de entrada. Você pode incluir o seguinte código:

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

inputGroupsPushPredicateTo

Retorna uma lista de grupos que podem receber a lógica de filtro. A transformação Java tem um grupo de entrada. Não modifique essa função para a transformação 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;
    }
}
```

Otimização de Envio com a Transformação Java

Você poderá habilitar uma transformação Java ativa para otimização de envio se ela não tiver efeitos colaterais e se a otimização não afetar os resultados do mapeamento.

Ao configurar a otimização de envio para a transformação Java, você define uma forma para essa transformação Java armazenar a condição de filtro que ela recebe do otimizador. Adicionar um código que examine a condição de filtro. Se a transformação Java puder absorver a lógica de filtro, a transformação Java transmitirá uma condição "true" de volta ao otimizador. O otimizador remove a transformação de Filtro do mapeamento otimizado.

Ao configurar a transformação Java, você grava o código que armazena a condição de filtro como metadados de transformação durante a otimização. Você também grava o código para recuperar a condição de filtro em tempo de execução e para descartar as linhas de acordo com a lógica de filtro.

Ao definir a transformação Java, você adiciona o código para otimização de envio na guia **Interfaces do Otimizador** da transformação Java. Para acessar os trechos de código para otimização de envio, escolha FilterPushdownOptimization no navegador da guia **Interfaces do Otimizador** da transformação.

A ferramenta Developer exibe trechos de código para habilitar a otimização de envio e para receber a condição de filtro do otimizador. Atualize os trechos de código para habilitar a otimização e para salvar a lógica de filtro como metadados de transformação.

isFilterSupported

Retorna "true" para habilitar a otimização de envio. Retorna "false" para desabilitar a otimização de envio. Altere a função para retornar "true" de forma a habilitar a otimização de envio.

```
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

Recebe a condição de filtro do otimizador.

Adicione um código para examinar o filtro e determinar se a lógica de filtro pode ser usada na transformação. Se a transformação puder absorver o filtro, use o método a seguir para armazenar a condição de filtro como metadados de transformação:

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

A chave é um identificador para os metadados. Você pode definir qualquer string como uma chave. Os dados são os dados que você deseja armazenar para determinar quais linhas devem ser descartadas em tempo de execução. Por exemplo, os dados podem ser a condição de filtro que a transformação Java recebe do otimizador.

```
public ResultAndMessage pushFilter(InfraExpression 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");
}
```

Otimização de Transformações de Associador

Transformações de Associador podem diminuir o desempenho, pois precisam de espaço adicional em tempo de execução para manterem resultados intermediários.

Considere as seguintes soluções para afunilamentos de transformações de Associador:

Designe a origem mestra como a origem com menos valores de chave duplicados.

Quando o Serviço de Integração de Dados processa uma transformação de Associador classificada, ele armazena em cache linhas de cem chaves exclusivas por vez. Se a origem mestra contiver muitas linhas com o mesmo valor de chave, o Serviço de Integração de Dados precisará armazenar mais linhas em cache, que poderá diminuir o desempenho.

Designe a origem mestra como a origem com menos linhas.

A transformação de Associador compara cada linha da origem de detalhes com a origem mestra. Quanto menos linhas houver na origem mestra, menor será o número de iterações da comparação de associações. Isso acelerará o processo de associação.

Quando possível, faça associações em um banco de dados.

Realizar uma associação em um banco de dados é mais rápido do que fazê-lo durante a execução do mapeamento. O tipo de associação de banco de dados usado pode afetar o desempenho. As associações normais são mais rápidas do que as associações externas e geram menos linhas. Em alguns casos, não é possível fazer a associação no banco de dados, como a associação de tabelas de dois bancos de dados diferentes ou de sistemas de arquivos simples.

Quando possível, associe dados classificados.

Configure a transformação de Associador para usar entrada classificada. O Serviço de Integração de Dados aumenta o desempenho, minimizando a entrada e a saída de disco. O maior aumento no desempenho ocorre quando você trabalha com grandes conjuntos de dados. No caso de uma transformação de Associador não classificada, designe a origem com menos linhas como a origem mestra.

Otimizar a condição de associação.

O Serviço de Integração de Dados tenta diminuir o tamanho do conjunto de dados de um operando de associação lendo as linhas a partir do grupo menor, localizando as linhas correspondentes no grupo maior e depois realizando a operação de associação. A diminuição do tamanho do conjunto de dados melhora o desempenho do mapeamento porque o Serviço de Integração de Dados deixa de ler linhas desnecessárias da origem de grupo maior. O Serviço de Integração de Dados move a condição de associação para a origem de grupo maior e lê apenas as linhas que correspondem ao grupo menor.

Usar o método de otimização com semi-associação.

Use o método de otimização com semi-associação para melhorar o desempenho do mapeamento quando um grupo de entrada possui muito mais linhas do que o outro e quando o grupo maior tem muitas linhas sem correspondência no grupo menor com base na condição de associação.

Otimização de Transformações de Pesquisa

Transformações de Pesquisa podem diminuir o desempenho dependendo do tipo de cache de pesquisa e das condições de pesquisa.

Considere as seguintes soluções para afunilamentos de transformações de Pesquisa:

Usar o driver de banco de dados ideal.

O Serviço de Integração de Dados consegue se conectar a uma tabela de pesquisa usando um driver de banco de dados nativo ou um driver ODBC. Drivers de banco de dados nativos oferecem um melhor desempenho de mapeamento em comparação a drivers ODBC.

Tabelas de pesquisa de cache para pesquisas de arquivo relacional ou simples.

Para melhorar o desempenho da pesquisa de origens de arquivo simples ou relacional, ative o cache de pesquisa na transformação. Quando você ativa o cache, o Serviço de Integração de Dados armazena a tabela de pesquisa no cache. Quando você executa o mapeamento, o Serviço de Integração de Dados consulta o cache de pesquisa em vez da tabela de pesquisa. Quando essa opção não está ativada, o Serviço de Integração de Dados consulta a tabela de pesquisa linha por linha.

O resultado do processamento e da consulta da pesquisa é o mesmo, independentemente de você armazenar ou não a tabela de pesquisa no cache. No entanto, o uso de um cache de pesquisa pode aumentar o desempenho do mapeamento em tabelas de pesquisa menores. Em geral, convém armazenar em cache as tabelas de pesquisa que precisam de menos de 300 MB.

Cache de tabelas de pesquisa para pesquisas de objeto de dados lógicos.

Para melhorar o desempenho da pesquisa em um objeto de dados lógicos, você pode ativar o cache do objeto de dados no Serviço de Integração de Dados. Quando você ativa o cache do objeto de dados, o Serviço de Integração de Dados armazena o objeto de dados lógicos em cache. Para ativar o cache do objeto de dados, você deve implantar o mapeamento em um aplicativo, ativar o cache do objeto de dados lógicos e executar o mapeamento com o comando `infacmd ms runmapping`. Quando você executa o mapeamento, o Serviço de Integração de Dados consulta o cache de objeto de dados em vez do objeto de dados lógicos.

Se você executar o mapeamento da Developer tool, a transformação de Pesquisa consultará o objeto de dados lógicos linha por linha.

Usar o tipo de cache adequado.

Use os seguintes tipos de caches para melhorar o desempenho:

- Cache compartilhado. Você pode compartilhar o cache de pesquisa entre várias transformações. Você pode compartilhar um cache não nomeado entre transformações no mesmo mapeamento. Você pode compartilhar um cache nomeado entre transformações no mesmo mapeamento ou em mapeamentos diferentes.
- Cache persistente. Para salvar e reutilizar os arquivos de cache, você pode configurar a transformação para usar um cache persistente. Use esse recurso quando você souber que a tabela de pesquisa não é alterada entre execuções de mapeamento. O uso de um cache persistente pode melhorar o desempenho, pois o Serviço de Integração de Dados cria o cache de memória a partir dos arquivos de cache, e não a partir do banco de dados.

Ativar caches simultâneos.

Quando o Serviço de Integração de Dados processa mapeamentos que contêm transformações de Pesquisa, ele cria um cache na memória ao processar a primeira linha de dados em uma transformação de Pesquisa armazenada em cache. Se houver várias transformações de Pesquisa em um mapeamento, o Serviço de Integração de Dados criará os caches sequencialmente quando a primeira linha de dados for processada pela transformação de Pesquisa. Isso deixa mais lento o processamento de transformações de Pesquisa.

Você pode ativar caches simultâneos para melhorar o desempenho. Quando o número de pipelines simultâneos adicionais está definido como um ou mais, o Serviço de Integração de Dados cria caches simultaneamente, e não sequencialmente. O desempenho melhora significativamente quando os mapeamentos contêm um número de transformações ativas que podem demorar para serem concluídas, como transformações de Agregador, Associador ou Classificador. Quando você ativa vários

pipelines simultâneos, o Serviço de Integração de Dados deixa de aguardar a conclusão de mapeamentos ativos antes de criar o cache. Outras transformações de Pesquisa no pipeline também criam caches simultaneamente.

Otimizar correspondências de condições de pesquisa.

Quando a transformação de Pesquisa corresponde aos dados do cache de pesquisa com a condição de pesquisa, ela classifica e organiza os dados para determinar o primeiro e o último valor correspondente. É possível configurar a transformação para retornar qualquer valor que corresponda à condição da pesquisa. Ao configurar a transformação de Pesquisa para retornar qualquer valor correspondente, a transformação retorna o primeiro valor que corresponde à condição de pesquisa. Ele não indexa todas as portas como faz quando você configura a transformação para retornar o primeiro ou o último valor correspondente.

Quando você usa um valor correspondente, o desempenho pode melhorar porque a transformação não é indexada em todas as portas, o que pode reduzir o desempenho.

Reduzir o número de linhas armazenadas em cache.

Você pode reduzir o tamanho das linhas incluídas no cache para melhorar o desempenho. Use a opção Substituição SQL de Pesquisa para adicionar uma cláusula WHERE à instrução SQL padrão. Ao adicionar uma cláusula WHERE a uma transformação de pesquisa usando um cache dinâmico, use uma transformação de Filtro antes da transformação de Pesquisa para transferir linhas ao cache dinâmico que correspondem a essa cláusula WHERE.

Substituir a instrução ORDER BY.

Por padrão, o Serviço de Integração de Dados gera uma instrução ORDER BY para uma pesquisa armazenada em cache. A instrução ORDER BY contém todas as portas de pesquisa. Para melhorar o desempenho, remova a instrução ORDER BY padrão e insira uma ORDER BY de substituição com menos colunas.

O Serviço de Integração de Dados sempre gera uma instrução ORDER BY, mesmo quando você insere outra na substituição. Coloque dois hífenes '--' depois da substituição ORDER BY para remover a instrução ORDER BY gerada.

Por exemplo, uma transformação de Pesquisa usa a seguinte condição de pesquisa:

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

A transformação de Pesquisa inclui três portas de pesquisa usadas no mapeamento, ITEM_ID, ITEM_NAME e PRICE. Quando você inserir a instrução ORDER BY, insira as colunas na mesma ordem que as portas na condição de pesquisa. Você também deve colocar todas as palavras reservadas do banco de dados entre aspas.

Insira a seguinte consulta de pesquisa na substituição SQL da pesquisa:

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

Usar uma máquina com mais memória.

Para aumentar o desempenho do mapeamento, execute o mapeamento em um nó do Serviço de Integração de Dados com uma grande quantidade de memória. Aumente os tamanhos do cache de dados e índices para o limite máximo, sem sobrecarregar a máquina. Se o nó do Serviço de Integração de Dados tiver memória suficiente, aumente o cache para que ele possa conter todos os dados na memória sem paginação no disco.

Otimizar a condição de pesquisa.

Se você incluir mais de uma condição de pesquisa, insira-as na seguinte ordem para otimizar o desempenho de pesquisa:

- Igual a (=)
- Menor que (<), maior que (>), menor que ou igual a (<=), maior que ou igual a (>=)
- Não é igual a (!=)

Filtrar linhas de pesquisa.

Para melhorar o desempenho, crie uma condição de filtro para reduzir o número de linhas de pesquisa recuperadas da origem quando o cache de pesquisa é criado.

Indexar a tabela de pesquisa.

O Serviço de Integração de Dados precisa consultar, classificar e comparar valores nas colunas de condição de pesquisa. O índice precisa incluir todas as colunas usadas em uma condição de pesquisa.

É possível melhorar o desempenho para os seguintes tipos de pesquisas:

- Pesquisas em cache. Para melhorar o desempenho, indexe as colunas na instrução ORDER BY de pesquisa. O arquivo de log de mapeamento contém a instrução ORDER BY.
- Pesquisas fora de cache. Para melhorar o desempenho, indexe as colunas na condição de pesquisa. O Serviço de Integração de Dados emite uma instrução SELECT para cada linha que passa pela transformação de Pesquisa.

Otimizar várias pesquisas.

Se um mapeamento contiver várias pesquisas, mesmo com o armazenamento em cache ativado e com memória de heap suficiente, as pesquisas poderão reduzir o desempenho. Ajuste as transformações de Pesquisa que consultam as maiores quantidades de dados para melhorar o desempenho geral.

Se a tabela de consultas estiver no mesmo banco de dados que a tabela de origem no seu mapeamento, e o armazenamento em cache não for viável, associe as tabelas no banco de dados de origem em vez de usar uma transformação de Pesquisa.

Otimização de Transformações de Classificador

Transformações de Classificador podem diminuir o desempenho quando a RAM física no nó do Data Integration Service não tem memória suficiente alocada para classificar dados.

Considere a seguinte solução para afunilamentos de transformações de Classificador:

Alocar memória suficiente.

Para obter o desempenho ideal, configure o tamanho do cache do Classificador com um valor menor que ou igual à quantidade de RAM física disponível no nó do Data Integration Service. Aloque pelo menos 16 MB de memória física para classificar dados usando a transformação de Classificador. Por padrão, o tamanho do cache do Classificador está definido como 16.777.216 bytes. Se o Data Integration Service não puder alocar memória suficiente para classificar dados, o mapeamento falhará.

Se a quantidade de dados de entrada for maior que o tamanho do cache do Classificador, o Data Integration Service armazenará temporariamente os dados no diretório de trabalho da transformação de Classificador. Quando armazena dados no diretório de trabalho, o Serviço de Integração de Dados requer que o espaço em disco seja pelo menos duas vezes maior que a quantidade de dados recebidos.

Otimização de Transformações SQL

Cada vez que o Serviço de Integração de Dados processa uma nova consulta em um mapeamento, ele chama uma função denominada SQLPrepare para criar um procedimento SQL e transmiti-lo ao banco de dados. Quando a consulta é alterada para cada linha de entrada, ela pode diminuir o desempenho.

Considere as seguintes soluções para afunilamentos de transformações SQL:

Não usar instruções de transação em uma consulta de transformação SQL.

Quando uma consulta SQL contém instruções de consulta de confirmação e reversão, o Serviço de Integração de Dados precisa recriar o procedimento SQL depois de cada confirmação ou reversão. Para otimizar o desempenho, não use instruções de transação em uma consulta de transformação SQL.

Habilitar métodos de otimização com seleção antecipada ou de otimização de filtros de envio com a transformação SQL.

Para melhorar o desempenho, considere habilitar a seleção antecipada ou o método de otimização de envio com a transformação SQL.

Otimização de Seleção Antecipada com a Transformação SQL

O Data Integration Service pode realizar a otimização de seleção antecipada com uma transformação SQL quando a condição de filtro faz referência somente portas de passagem, e a transformação SQL não apresenta efeitos colaterais.

A transformação SQL apresenta efeitos colaterais nas seguintes circunstâncias:

- A consulta SQL atualiza um banco de dados. A consulta SQL contém uma instrução como CREATE, DROP, INSERT, UPDATE, GRANT ou REVOKE.
- A transformação retorna linhas NULL para instruções SELECT que não retornam resultados. As linhas podem conter valores de porta de passagem, informações de erros SQL ou o campo NumRowsAffected.

Habilitando a Otimização de Seleção Antecipada com a Transformação SQL

Habilite a otimização com seleção antecipada na transformação SQL quando esta última não apresenta efeitos colaterais.

1. Habilite a opção **Retornar Somente Saída de Banco de Dados** nas **Propriedades Avançadas** da transformação SQL.
2. Desmarque a opção **Tem Efeitos Colaterais** nas **Propriedades Avançadas** da transformação.
3. Se a transformação tiver uma porta **NumAffectedRows**, remova essa porta.

Otimização de Envio com a Transformação SQL

Com a otimização de envio, o Serviço de Integração de Dados envia a lógica de filtro de uma transformação de Filtro no mapeamento para a consulta na transformação SQL.

Use as seguintes regras e diretrizes quando você habilitar a otimização de envio com a transformação SQL:

- A consulta de transformação SQL deve conter somente instruções SELECT.
- A consulta SQL da transformação deve ser uma subconsulta válida.
- A condição de filtro não pode fazer referência aos campos Erro SQL ou NumRowsAffected.
- Os nomes das portas de saída devem corresponder aos nomes das colunas na instrução SQL SELECT. Quando você faz referência a uma porta de saída em uma condição de filtro, o Serviço de Integração de Dados envia o nome da porta de entrada correspondente à consulta SQL. Você pode adicionar aliases ao

SQL quando as colunas na consulta não correspondem aos nomes das portas de saída. Por exemplo, `SELECT mycolname1 AS portname1, mycolname2 AS portname2.`

- A transformação não pode ter efeitos colaterais.

Otimização de Envio com a Transformação SQL - Exemplo

Uma transformação SQL recupera pedidos por ID de cliente. Uma transformação de Filtro que aparece após a transformação SQL retorna apenas as linhas em que a quantidade de pedidos é maior que 1000.

O Serviço de Integração de Dados envia o seguinte filtro em uma instrução SELECT na transformação SQL:

```
orderAmount > 1000
```

Cada instrução na consulta SQL se torna uma subconsulta separada da instrução SELECT que contém o filtro.

A seguinte instrução de consulta mostra a instrução de consulta original como uma subconsulta na instrução SELECT:

```
SELECT <customerID>, <orderAmount>, ... FROM (instruções de consulta originais) ALIAS WHERE  
<orderAmount> > 1000
```

Se a consulta SQL tiver várias instruções, cada uma será incluída em uma subconsulta separada. A subconsulta tem a mesma sintaxe, incluindo a cláusula WHERE.

As portas *customerID* e *orderAmount* são os nomes das portas de saída na transformação SQL. A subconsulta não inclui portas de passagem, o erro SQL ou as portas de estatísticas SQL. Se você enviar vários filtros para a transformação SQL, a cláusula WHERE conterá todos esses filtros.

Habilitando a Otimização de Envio com a Transformação SQL

Habilite a otimização de envio configurando propriedades na guia **Propriedades Avançadas** da transformação SQL.

1. Desmarque **Tem Efeitos Colaterais**.
2. Habilite **Retornar Somente Saída de Banco de Dados**.
3. Defina **Número de Linhas de Saída Máximas** como zero.
4. Habilite a otimização de envio.

Cache de Transformações

Quando você executa um mapeamento que usa uma transformação de Agregador, de Associador, de Pesquisa, de Classificação ou de Classificador, o Serviço de Integração de Dados cria caches na memória para processar a transformação. Se o Serviço de Integração de Dados precisar de mais espaço, ele armazenará os valores de overflow nos arquivos de cache no disco.

Considere a seguinte solução para afunilamentos de cache de transformação:

Configure as transformações para alocar espaço suficiente para armazenar o cache na memória.

Para melhorar o tempo de processamento da transformação de Agregador, de Associador, de Pesquisa, de Classificação ou de Classificador, configure essas transformações para alocar espaço suficiente para armazenar o cache na memória. Quando você configura a quantidade de memória cache de forma

que ela seja igual ou superior à quantidade necessária para armazenar os dados e o índice em cache, o desempenho pode ser melhorado reduzindo a sobrecarga de E/S do sistema. Quando o Serviço de Integração de Dados grava arquivos de cache no disco, o tempo de processamento aumenta devido à sobrecarga de E/S do sistema.

Por padrão, o Serviço de Integração de Dados configura automaticamente os requisitos de memória cache no tempo de execução. Após executar um mapeamento no modo de cache automático, você poderá ajustar os tamanhos de cache das transformações. Analise as estatísticas de transformação no log de mapeamento para determinar os tamanhos de cache necessários para processar as transformações na memória. Ao configurar o tamanho do cache para usar o valor especificado no log de mapeamento, você pode garantir que nenhuma memória alocada seja desperdiçada. No entanto, o tamanho do cache ideal varia com base no tamanho da fonte de dados. Analise os logs de mapeamento depois que o mapeamento subsequente for executado para monitorar as alterações no tamanho do cache. Se você configurar um tamanho de cache específico para uma transformação reutilizável, verifique se o tamanho do cache é ideal para cada uso da transformação em um mapeamento.

Eliminação de Erros de Transformação

Em grandes números, erros de transformação reduzem o desempenho do Data Integration Service. Com cada erro de transformação o Data Integration Service é pausado para determinar a causa do erro e para remover do fluxo de dados a linha causadora do erro. Em geral, o Data Integration Service geralmente grava a linha no arquivo de log de mapeamento dos logs do Data Integration Service.

Considere as seguintes soluções para afunilamentos de erros de transformação:

Consultar o arquivo de log de mapeamento para verificar onde os erros de transformação ocorrem e avaliar essas restrições de transformação.

Erros de transformação ocorrem quando o Data Integration Service encontra erros de conversão, lógica de mapeamento conflitante e qualquer condição configurada como um erro, como uma entrada inválida. Consulte o arquivo de log de mapeamento para verificar onde os erros de transformação ocorrem. Se os erros se concentrarem em transformações específicas, avalie as restrições dessas transformações.

Configurar um nível de rastreamento mais baixo.

Se você precisar executar um mapeamento que gera muitos erros de transformação, será possível melhorar o desempenho configurando um nível de rastreamento inferior. No entanto, essa não é uma solução recomendada a longo prazo para erros de transformação.

Efeitos Colaterais de Transformações

Uma transformação apresenta efeitos colaterais quando retorna linhas e modifica um objeto ou quando interage com outros objetos ou funções. A transformação pode modificar um banco de dados, adicionar a um total, gerar uma exceção, escrever um email ou chamar outras funções com efeitos colaterais.

O Serviço de Integração de Dados identifica quais transformações apresentam efeitos colaterais antes de otimizar um mapeamento. O Serviço de Integração de Dados pressupõe que uma transformação apresenta efeitos colaterais quando ele não pode determinar se a transformação apresenta efeitos colaterais.

Transformações com efeitos colaterais limitam quando o Serviço de Integração de Dados pode otimizar o mapeamento. A otimização de seleção antecipada, de remoção de ramificação, de predicado global e de

envio alterarão os resultados do mapeamento se o Serviço de Integração de Dados as aplicar a uma transformação que apresenta efeitos colaterais. Com a otimização de envio e a otimização com seleção antecipada, a lógica de filtro é movida de uma transformação de Filtro o mais próximo possível da origem. Se o filtro ocorrer antes da função de efeito colateral, os resultados do mapeamento mudarão.

Por exemplo, uma transformação recebe um ID de cliente e retorna linhas que contêm informações de pedidos. A transformação também grava os pedidos em um arquivo. Se o Serviço de Integração de Dados aplicar uma otimização de filtro antes de gravar pedidos no arquivo, esse arquivo receberá menos linhas do que quando o filtro ocorrer posteriormente no mapeamento. O efeito colateral da transformação é a função de gravar os registros de pedidos em um arquivo.

As seguintes transformações apresentam efeitos colaterais:

- Transformação SQL, transformação de Consumidor de Serviço da Web e transformação Java, a menos que a propriedade de efeitos colaterais esteja desabilitada.
- Transformações que chamam uma função `ABORT()` ou `ERROR()`, enviam e-mail ou chamam um procedimento armazenado.
- Transformações que gravam em arquivos ou bancos de dados.
- Transformações que mantêm uma contagem através de uma porta variável. Por exemplo, `COUNT=COUNT+1`.

A transformação SQL, a transformação de Consumidor de Serviço da Web e a transformação Java apresentam efeitos colaterais por padrão. Se você configurar a transformação para processar linhas sem efeitos colaterais, poderá desabilitar a propriedade **Tem Efeitos Colaterais** em **Propriedades Avançadas**. Se a transformação não tiver efeitos colaterais, você poderá habilitar a otimização configurando propriedades adicionais nessas transformações.

Otimização de Transformações de Consumidor de Serviço da Web

A transformação de Consumidor de Serviço da Web pode diminuir o desempenho quando um mapeamento chama o serviço da Web várias vezes.

Considere a seguinte solução para afunilamentos de transformações de Consumidor de Serviço da Web:

Configurar a transformação de Consumidor de Serviço da Web para usar a autenticação por cookies.

O servidor remoto de serviços da Web rastreia os usuários consumidores de serviços da Web com base nos cookies. Você pode melhorar o desempenho quando um mapeamento chama um serviço da Web várias vezes.

Quando a porta para cookies é projetada para uma mensagem de solicitação de serviço da Web, o provedor do serviço da Web retorna um valor de cookie na mensagem de resposta. Você pode transmitir o valor do cookie para outra transformação mais abaixo no mapeamento ou pode salvar o valor do cookie em um arquivo. Ao salvar o valor do cookie em um arquivo, você pode configurar esse cookie como entrada para a transformação de Consumidor de Serviço da Web. É possível projetar a porta de saída de cookies para qualquer um dos grupos de saída de transformações de Consumidor de Serviço da Web.

Habilitar métodos de otimização com seleção antecipada ou de otimização de filtros de envio com a transformação de Consumidor de Serviço da Web.

Para aumentar o desempenho, o Serviço de Integração de Dados pode aplicar o método de seleção antecipada ou de otimização de envio com a transformação de Consumidor de Serviço da Web. Para

aplicar a otimização com seleção antecipada, o serviço da Web não pode apresentar efeitos colaterais e não pode tratar falhas como erros. Para aplicar a otimização de envio, o serviço da Web não pode apresentar efeitos colaterais, não pode tratar falhas como erros, e a condição de filtro deve fazer referência a portas de passagem.

O serviço da Web apresenta efeitos colaterais quando realiza outras funções além de retornar uma resposta à transformação de Consumidor de Serviço da Web. O serviço da Web apresenta efeitos colaterais quando modifica um banco de dados, grava em um arquivo, escreve emails, atualiza uma contagem ou chama outros serviços da Web com efeitos colaterais.

Otimização de Seleção Antecipada com a Transformação de Consumidor de Serviço da Web

Quando o Data Integration Service aplica o método de otimização com seleção antecipada à transformação de Consumidor de Serviço da Web, ele aproxima da origem as condições de filtro antes da transformação de Consumidor de Serviço da Web no mapeamento.

Habilitando a Otimização de Seleção Antecipada com a Transformação de Consumidor de Serviço da Web

Habilite a otimização com seleção antecipada para a transformação de Consumidor de Serviço da Web quando essa transformação não apresenta efeitos colaterais e não trata falhas como erros.

1. Abra a exibição **Propriedades Avançadas** da transformação de Consumidor de Serviço da Web.
2. Desmarque **Tratar Falha como Erro**.
3. Desmarque **Tem Efeitos Colaterais**.

Otimização de Envio com a Transformação de Consumidor de Serviço da Web

Você pode configurar a otimização de envio com a transformação de Consumidor de Serviço da Web quando essa transformação está em uma tabela virtual de um serviço de dados SQL.

O mapeamento chama o serviço da Web para recuperar um conjunto de dados ou um subconjunto dos dados com base nas instruções na consulta SQL do usuário final. A consulta SQL do usuário final contém uma condição de filtro opcional.

Com a otimização de envio, a transformação de Consumidor de Serviço da Web recebe o valor do filtro em uma porta de filtro. A porta de filtro é uma porta de entrada não conectada que você identifica como uma porta de filtro quando configura a otimização de envio. A porta de filtro tem um valor padrão que garante que o serviço da Web retorne todas as linhas quando a consulta do usuário final não contém filtros. A porta de filtro não é uma porta de passagem.

Nota: O campo de filtro deve fazer parte do grupo raiz na solicitação de serviço da Web.

Ao configurar uma porta de filtro, você identifica uma porta de saída na transformação de Consumidor de Serviço da Web que recebe os dados de colunas da resposta do serviço da Web. Por exemplo, se a porta de filtro for uma porta de entrada denominada EmployeeID, a porta de saída da resposta poderá ser uma porta denominada EmployeeNum. A ferramenta Developer precisa associar a porta de filtro de entrada a uma porta de saída para enviar a lógica de filtro da leitura da tabela virtual para a solicitação de Consumidor de Serviço da Web. As portas de entrada para uma solicitação de serviço da Web são normalmente diferentes das portas de saída da resposta de serviço da Web.

O campo de filtro não pode ser uma porta de passagem. Quando você configura uma porta de filtro, o valor padrão da porta se transforma no valor da condição de filtro e, portanto, o valor da porta de saída de passagem é alterado. Um filtro baseado na porta de passagem de saída retorna resultados inesperados.

Você pode enviar várias expressões de filtro para a transformação de Consumidor de Serviço da Web. Cada condição de filtro deve ter o seguinte formato:

<Field> = <Constant>

As condições de filtro devem ser associadas por AND. Não é possível associar as condições com um OR.

Otimização de Envio com a Transformação de Consumidor de Serviço da Web - Exemplo

Um serviço de dados SQL retorna pedidos para todos os clientes ou retorna pedidos para um cliente específico com base na consulta SQL que ele recebe do usuário.

O serviço de dados contém um objeto de dados lógicos com os seguintes componentes:

Tabela Customer

Uma tabela de banco de dados Oracle que contém informações de clientes.

Transformação de Consumidor de Serviço da Web

Uma transformação que chama um serviço da Web para recuperar os pedidos mais recentes para clientes. A transformação de Consumidor de Serviço da Web tem portas de entrada para customerID e orderNum. A transformação tem portas de passagem que contêm dados de clientes que ela recebe da tabela Customer. A porta orderNum é a porta de filtro e não está conectada. orderNum tem o valor padrão "*". Quando o serviço da Web recebe esse valor na solicitação de serviço da Web, ele retorna todos os pedidos.

Tabela virtual Orders

Uma tabela virtual que recebe os dados de pedidos e clientes do serviço da Web. O usuário final consulta essa tabela. A tabela Orders contém uma coluna "customer", uma coluna orderID e dados de clientes e pedidos.

O usuário final transmite a seguinte consulta SQL ao serviço de dados SQL:

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

O Data Integration Service divide a consulta para otimizar o mapeamento. O Data Integration Service usa a otimização com seleção antecipada e move a lógica do filtro, `customer = 23`, para a leitura da tabela Customer. O Data Integration Service usa a otimização de envio e envia a lógica de filtro, `orderID = 56`, para a porta de filtro da transformação de Consumidor de Serviço da Web. A transformação de Consumidor de Serviço da Web recupera ordersID 56 para customer 23.

Habilitando a Otimização de Envio com a Transformação de Consumidor de Serviço da Web

Habilite a otimização de envio para a transformação de Consumidor de Serviço da Web quando essa transformação não apresenta efeitos colaterais e não trata falhas como erros.

1. Abra a exibição **Propriedades Avançadas** da transformação de Consumidor de Serviço da Web.
2. Desmarque **Tratar Falha como Erro**.
3. Desmarque **Tem Efeitos Colaterais**.
4. Clique no botão **Abrir** na propriedade **Otimização de Envio**.

5. Escolha o nome da porta de filtro na caixa de diálogo Entrada Otimizada. Você pode escolher várias portas de filtro.
6. Clique na coluna **Saída**.
7. Para cada porta de filtro, escolha a porta de saída que contém a coluna filtrada na resposta do serviço da Web.
8. Insira um valor padrão para cada porta de filtro.

Nota: Não é possível configurar um valor padrão para uma porta de Consumidor de Serviço da Web a menos que ela seja uma porta de filtro.

CAPÍTULO 5

Otimização de Mapeamentos

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Otimização de Mapeamentos, 41](#)
- [Métodos de Otimização, 42](#)
- [Otimização de Empilhamento, 49](#)
- [Leitura de Passagem Única, 51](#)
- [Otimização de Filtros, 52](#)
- [Otimização de Conversão de Tipos de Dados, 52](#)
- [Rastreamento de Erros, 53](#)

Visão Geral da Otimização de Mapeamentos

Otimize mapeamentos para permitir que o Serviço de Integração de Dados transforme e mova dados de maneira eficiente. A otimização em nível de mapeamento pode demorar a ser implementada, mas pode melhorar significativamente o desempenho de mapeamento.

As tarefas de otimização se aplicam a mapeamentos regulares, mapeamentos de leitura e gravação de objetos de dados lógicos, mapeamentos de tabela virtual e mapeamentos de operação. Enfatize a otimização em nível de mapeamento depois de otimizar os destinos e as origens.

Para otimizar um mapeamento, você pode realizar as seguintes tarefas:

- Configure o mapeamento de forma que o mínimo de transformações e expressões faça a maior quantidade de trabalho possível.
- Exclua os links desnecessários entre as transformações para minimizar a quantidade de dados movidos.
- Escolha um nível do otimizador que determine quais métodos de otimização o Serviço de Integração de Dados pode aplicar ao mapeamento. Quando o Serviço de Integração de Dados otimiza um mapeamento, ele tenta reduzir a quantidade de dados a serem processados. Por exemplo, o Serviço de Integração de Dados pode usar a otimização com seleção antecipada para aproximar um filtro da origem. Ele pode usar o método de otimização baseada em custos para alterar a ordem de processamento de associações.
- Selecione um tipo de empilhamento para ativar o Serviço de Integração de Dados para determinar se ele pode empilhar a lógica de transformação completa ou parcial para o banco de dados de origem.
- Configure o armazenamento em cache do objetos de dados para permitir que o Serviço de Integração de Dados armazene objetos de dados lógicos em cache e acesse objetos de dados lógicos pré-criados quando executar um mapeamento. Por padrão, o Serviço de Integração de Dados extrai dados de origem e cria os objetos de dados necessários quando executa um mapeamento. O desempenho do

mapeamento melhora quando o Serviço de Integração de Dados pode acessar objetos de dados pré-criados.

- Indique se a transformação SQL, a transformação de Consumidor de Serviço da Web e a transformação Java não têm efeitos colaterais quando você configurá-las. Algumas transformações têm efeitos colaterais que restringem a otimização. Por exemplo, uma transformação pode ter um efeito colateral quando grava em um arquivo ou banco de dados, adiciona a uma contagem, gera uma exceção ou escreve um e-mail. Na maioria dos casos, o Serviço de Integração de Dados identifica qual transformação tem efeitos colaterais que restringem a otimização.

TÓPICOS RELACIONADOS:

- [“Armazenamento em Cache do Objeto de Dados” na página 65](#)

Métodos de Otimização

O Serviço de Integração de Dados aplica métodos de otimização para reduzir o número de linhas a serem processadas no mapeamento. Você pode configurar o nível do otimizador do mapeamento para limitar os métodos de otimização aplicados pelo Serviço de Integração de Dados.

O Serviço de Integração de Dados pode aplicar os seguintes métodos de otimização:

- Otimização de empilhamento
- Otimização de projeção antecipada
- Otimização de seleção antecipada
- Otimização de remoção de ramificação
- Otimização de envio
- Otimização de predicado
- Otimização de predicado global
- Otimização baseada em custos
- Otimização dataship-join
- Otimização de semiassociação

O Serviço de Integração de Dados pode aplicar vários métodos de otimização a um mapeamento ao mesmo tempo. Por exemplo, o Serviço de Integração de Dados aplica os métodos de otimização de projeção antecipada, otimização de predicado, otimização de predicado global, otimização de remoção de ramificação e otimização de seleção antecipada ou otimização de envio quando você seleciona o nível do otimizador normal.

Níveis do Otimizador

O Serviço de Integração de Dados otimiza os mapeamentos com base no nível de otimizador que você configurar. Configure o nível do otimizador quando você desejar que o mapeamento utilize um nível do otimizador diferente do normal. Por padrão, cada mapeamento usa o nível de otimizador normal.

Você pode escolher um dos seguintes níveis de otimizador:

Nenhum

O Serviço de Integração de Dados não aplica nenhuma otimização.

Mínima

O Serviço de Integração de Dados aplica o método de otimização de projeção antecipada.

Normal

O Serviço de Integração de Dados aplica os métodos de otimização de projeção antecipada, seleção antecipada, remoção de ramificação, envio, predicado global e predicado. Normal é o nível de otimização padrão.

Completo

O Serviço de Integração de Dados aplica os métodos de otimização baseada em custos, de projeção antecipada, de seleção antecipada, remoção de ramificação, predicado, envio, semiassociação e dataship-join.

O Serviço de Integração de Dados aplica o nível do otimizador normal quando você executa um mapeamento no menu **Executar** ou no editor de mapeamento na Developer tool. Quando você executa o mapeamento no menu **Executar**, o Serviço de Integração de Dados aplica o nível do otimizador na configuração de mapeamento. Quando o mapeamento é executado da linha de comando, o Serviço de Integração de Dados aplica o nível de otimização das propriedades de implantação de mapeamento no aplicativo.

Nota: O Serviço de Integração de Dados não aplica o método de otimização de empilhamento com um nível do otimizador. Você pode configurar a otimização de empilhamento para um mapeamento nas propriedades de tempo de execução do mapeamento.

Otimização de Filtro

A otimização de filtro melhora o desempenho ao reduzir o número de linhas que passam pelo mapeamento. O Data Integration Service pode aplicar a otimização com seleção antecipada ou a otimização de envio.

Quando o Data Integration Service aplica um método de otimização de filtro, ele move um filtro o mais próximo possível da origem em um mapeamento. Se o Data Integration Service não conseguir mover um filtro antes de uma transformação em um mapeamento, talvez ele consiga enviar a lógica do filtro para uma transformação.

Método de Otimização de Projeção Antecipada

Quando o Serviço de Integração de Dados aplica o método de otimização de projeção antecipada, ele identifica portas não utilizadas e remove os links entre essas portas.

O método de otimização de projeção antecipada melhora o desempenho reduzindo a quantidade de dados que o Serviço de Integração de Dados movimenta entre as transformações. Quando o Serviço de Integração de Dados processa um mapeamento, ele movimenta os dados de todas as portas conectadas em um mapeamento de uma transformação para outra. Em mapeamentos grandes e complexos, ou em mapeamentos que usam mapplets aninhados, algumas portas podem não fornecer dados para o destino. O Serviço de Integração de Dados identifica as portas que não fornecem dados para o destino. Depois que o Serviço de Integração de Dados identifica portas não utilizadas, ele remove do mapeamento os links entre todas essas portas.

O Serviço de Integração de Dados não remove todos os links. Por exemplo, ele não remove os seguintes links:

- Links conectados a uma transformação que tem efeitos colaterais.
- Links conectados a transformações que chamam uma função ABORT() ou ERROR(), enviam e-mail ou chamam um procedimento armazenado.

Se o Serviço de Integração de Dados determinar que todas as portas em uma transformação estão fora de uso, ele removerá todos os links de transformação, exceto o link para a porta com o mínimo de dados. O Serviço de Integração de Dados não remove a transformação não utilizada do mapeamento.

A Developer tool habilita esse método de otimização por padrão.

Método de Otimização de Predicado

Quando o Data Integration Service aplica o método de otimização de predicado, ele examina as expressões de predicado geradas por um mapeamento. Ele determina se pode simplificar ou reescrever as expressões para melhorar o desempenho do mapeamento.

Quando o Data Integration Service executa um mapeamento, ele gera consultas nas origens do mapeamento e realiza operações nos resultados das consultas com base na lógica de mapeamento e nas transformações dentro desse mapeamento. As consultas e operações muitas vezes incluem expressões de predicado. Expressões de predicado representam as condições que os dados devem satisfazer. O filtro e as condições de associação em transformações de Filtro e de Associador são exemplos de expressões de predicado.

Com o método de otimização de predicado, o Data Integration Service também tenta aplicar expressões de predicado com a maior antecedência possível no mapeamento para melhorar o desempenho do mapeamento.

O Data Integration Service deduz relacionamentos a partir de expressões de predicado existentes e cria novas expressões de predicado. Por exemplo, um mapeamento contém uma transformação de Associador com a condição de associação "A=B" e uma transformação de filtro com a condição de filtro "A>5". O Data Integration Service pode adicionar "B>5" à condição de associação.

O Data Integration Service aplica o método de otimização de predicado com o método de otimização de seleção antecipada quando pode aplicar os dois métodos a um mapeamento. Por exemplo, quando o Data Integration Service cria novas condições de filtro usando o método de otimização de predicado, ele também tenta movê-las em direção ascendente no mapeamento usando o método de seleção antecipada. Aplicar ambos os métodos de otimização melhora o desempenho do mapeamento em comparação a aplicar um desses métodos isoladamente.

O Data Integration Service aplica o método de otimização de predicado quando essa aplicação melhora o desempenho. O Serviço de Integração de Dados não aplicará esse método se o aplicativo alterar os resultados ou diminuir o desempenho do mapeamento. O Serviço de Integração de Dados aplica esse método de otimização por padrão.

Regras e Diretrizes para Otimização de Predicado

Quando o Serviço de Integração de Dados regravava uma expressão de predicado, ele aplica uma lógica matemática à expressão para otimizá-la.

O Serviço de Integração de Dados pode executar qualquer uma ou todas as ações a seguir:

- Identifica variáveis equivalentes em expressões de predicado no mapeamento e gera expressões simplificadas com base nas equivalências.
- Identifica predicados redundantes em expressões de predicado no mapeamento e os remove.
- Extrai subexpressões de cláusulas disjuntivas e gera várias expressões simplificadas com base nas subexpressões.
- Normaliza uma expressão de predicado.
- Aplica expressões de predicado com a maior antecedência possível no mapeamento.

O Serviço de Integração de Dados pode não aplicar a otimização de predicado a um mapeamento quando esse mapeamento contém transformações com uma incompatibilidade de tipo de dados entre portas conectadas.

O Serviço de Integração de Dados pode não aplicar a otimização de predicado a uma transformação quando qualquer uma das seguintes condições é verdadeira:

- A transformação contém valores padrão explícitos para portas conectadas.
- A transformação tem efeitos colaterais.
- A transformação não permite que predicados sejam movidos. Por exemplo, uma transformação que tem efeitos colaterais pode ter essa restrição.

A ferramenta Developer permite o método de otimização de predicado por padrão.

Método de Otimização Baseada em Custos

Com a otimização baseada em custos, o Serviço de Integração de Dados avalia um mapeamento, gera mapeamentos semanticamente equivalentes e executa o mapeamento com o melhor desempenho possível. A otimização baseada em custos reduz o tempo de execução dos mapeamentos que realizam operações de associação inner-join e full-outer adjacentes.

Mapeamentos semanticamente equivalentes são mapeamentos que realizam funções idênticas e produzem os mesmos resultados. Para gerar mapeamentos semanticamente equivalentes, o Data Integration Service divide o mapeamento original em fragmentos. O Data Integration Service determina então quais fragmentos de mapeamento ele pode otimizar.

Durante a otimização, o Data Integration Service pode adicionar, remover ou reorganizar transformações dentro de um fragmento. O Data Integration Service verifica se os fragmentos otimizados produzem os mesmos resultados que os fragmentos originais e forma mapeamentos alternativos que usam esses fragmentos otimizados.

O Serviço de Integração de Dados também poderá aplicar uma associação de mesclagem classificada se determinar que o desempenho dela é melhor do que o desempenho da associação de loop aninhado. Uma associação de mesclagem classificada usa a ordem de classificação para organizar dois conjuntos de dados antes de realizar a integração. Uma associação de loop aninhado usa loops aninhados para associar dois conjuntos de dados. O Serviço de Integração de Dados poderá usar as informações de classificação nas origens ou criar uma transformação do Classificador se o custo da classificação dos dados for menor do que o processamento da associação de loop aninhado.

O Data Integration Service gera todos, ou quase todos, os mapeamentos que são semanticamente equivalentes ao mapeamento original. Ele usa estatísticas de criação de perfil ou estatísticas de banco de dados para calcular o custo do mapeamento original e de cada mapeamento alternativo. Em seguida, ele identifica o mapeamento que é executado com a maior rapidez. O Data Integration Service realiza uma verificação de validação no melhor mapeamento alternativo para garantir que ele seja válido e produza os mesmos resultados que o mapeamento original.

O Data Integration Service armazena no cache da memória o melhor mapeamento alternativo. Quando você executar um mapeamento, o Data Integration Service recuperará o mapeamento alternativo e o executará no lugar no mapeamento original.

A Developer tool não ativa esse método por padrão.

Método de Otimização Dataship-Join

O método de otimização dataship-join tenta localizar os conjuntos de dados menores ao lado dos conjuntos de dados maiores para reduzir o tempo de processamento da associação. O Serviço de Integração de Dados

tenta aplicar o método de otimização dataship-join quando há uma diferença significativa de tamanho entre duas tabelas.

Por exemplo, o Serviço de Integração de Dados pode aplicar o método de otimização dataship-join para associar uma tabela mestre que contém 10.000 linhas a uma tabela de detalhes que contém 1.000.000 linhas. Para realizar o dataship-join, o Serviço de Integração de Dados cria uma tabela de dados temporária no banco de dados que contém a tabela de detalhes maior. Em seguida, o Serviço de Integração de Dados copia a tabela mestre menor para uma tabela temporária e associa os dados na tabela temporária aos dados na tabela de detalhes maior. Depois que o Serviço de Integração de Dados realiza a operação de associação, a lógica da transformação do Associador é processada no banco de dados.

Antes de aplicar esse método de otimização, o Serviço de Integração de Dados realiza análises para determinar se a otimização dataship-join é possível e se vale a pena realizá-la. Se a análise determinar que esse método provavelmente melhorará o desempenho, o Serviço de Integração de Dados o aplicará ao mapeamento. Em seguida, o Serviço de Integração de Dados repete a análise do mapeamento para determinar se existem outras oportunidades para uma otimização dataship-join. Se adequado, ele realizará mais otimizações.

A Developer tool não ativa esse método por padrão.

Requisitos do Dataship-Join para Melhor Desempenho

O método de otimização dataship-join nem sempre melhora o desempenho. Os seguintes fatores afetam o desempenho do mapeamento na otimização dataship-join:

- A origem mestre da transformação de Associador deve ter um número de linhas significativamente menor que a origem dos detalhes.
- A origem detalhada deve ser muito grande para justificar a otimização. Se a origem detalhada não for grande o suficiente, o Serviço de Integração de Dados achará mais rápido ler todos os dados da origem mestre e detalhada sem aplicar o método de otimização dataship-join.

Regras e Diretrizes da Otimização Dataship-Join

O Serviço de Integração de Dados poderá aplicar a otimização dataship-join a uma transformação do Associador se a transformação atender aos seguintes requisitos:

- O tipo de associação deve ser normal, externa mestre ou externa detalhada.
- O pipeline detalhado deve ser proveniente de uma origem relacional.
- Se o mapeamento usar confirmações baseadas em destinos, o escopo da transformação do Associador deverá ser Todas as Entradas.
- O pipeline mestre e o pipeline detalhado não podem compartilhar transformações.
- O mapeamento não pode conter uma ramificação entre a origem de detalhes e a transformação de Associador.
- O Serviço de Integração de Dados não conseguirá aplicar o método de otimização dataship-join se o banco de dados que contém o lado de detalhes da associação for um banco de dados IBM DB2 sem suporte para codificação Unicode.

Método de Otimização de Semi-associação

O método de otimização de semi-associação tenta reduzir a quantidade de dados extraídos da origem modificando operações de associação no mapeamento.

O Serviço de Integração de Dados aplica o método de otimização de semiassociação a uma transformação do Associador quando um grupo de entrada tem muito mais linhas do que outro e quando o grupo maior tem

muitas linhas sem correspondência no grupo menor com base na condição de associação. O Data Integration Service tenta diminuir o tamanho do conjunto de dados de um operando de associação lendo as linhas a partir do grupo menor, localizando as linhas correspondentes no grupo maior e depois realizando a operação de associação. A diminuição do tamanho do conjunto de dados melhora o desempenho do mapeamento porque o Data Integration Service deixa de ler linhas desnecessárias da origem de grupo maior. O Data Integration Service move a condição de associação para a origem de grupo maior e lê apenas as linhas que correspondem ao grupo menor.

Antes de aplicar o método de otimização de semiassociação, o Serviço de Integração de Dados realiza análises para determinar se ela é possível e se vale a pena realizá-la. Se a análise determinar que esse método provavelmente melhorará o desempenho, o Serviço de Integração de Dados o aplicará ao mapeamento. Em seguida, o Data Integration Service repete a análise do mapeamento para determinar se existem oportunidades adicionais para uma otimização de semi-associação. Se adequado, ele realizará mais otimizações.

A Developer tool não ativa esse método por padrão.

Requisitos de Otimização com Semi-associação para Melhorar o Desempenho

O método de otimização com semi-associação nem sempre melhora o desempenho. Os seguintes fatores afetam o desempenho do mapeamento na otimização com semi-associação:

- A origem mestre da transformação de Associador deve ter um número de linhas significativamente menor que a origem dos detalhes.
- A origem dos detalhes deve ser grande o suficiente para justificar a otimização. Quando o Data Integration Service aplica a otimização com semi-associação, o método adiciona um certo tempo de sobrecarga ao processamento do mapeamento. Se a origem dos detalhes for pequena, o tempo necessário para aplicar o método de semi-associação poderá exceder o tempo necessário para processar todas as linhas na origem dos detalhes.
- O Data Integration Service deve ser capaz de obter estatísticas de número de linhas de origem para uma transformação de Associador para comparar de forma precisa os requisitos de tempo da operação de associação regular com os da operação de semi-associação.

Regras e Diretrizes para Otimização de Semi-associação

O Data Integration Service poderá aplicar a otimização com semi-associação a uma transformação de Associador se essa transformação atender aos seguintes requisitos:

- O tipo de associação deve ser normal, externa mestre ou externa detalhada. A transformação de Associador não pode realizar uma associação externa completa.
- O pipeline detalhado deve ser proveniente de uma origem relacional.
- A condição de associação deve ser uma condição válida de associação/classificação/mesclagem. Isto é, cada uma cláusula deve ser uma igualdade de uma porta mestre e uma porta detalhada. Se houver várias cláusulas, será necessário associá-las com AND.
- Se o mapeamento não usar confirmações baseadas em destinos, o escopo da transformação de Associador deverá ser Todas as Entradas.
- O pipeline mestre e o pipeline detalhado não podem compartilhar transformações.
- O mapeamento não pode conter uma ramificação entre a origem de detalhes e a transformação de Associador.

Método de Otimização de Seleção Antecipada

Quando o Serviço de Integração de Dados aplica o método de otimização com seleção antecipada, ele divide, move ou remove as transformações de Filtro em um mapeamento. Ele move os filtros para cima no mapeamento, aproximando-os da origem.

O Serviço de Integração de Dados pode dividir uma transformação de Filtro quando a condição de filtro é uma conjunção. Por exemplo, o Serviço de Integração de Dados pode dividir a condição de filtro "A>100 AND B<50" em duas condições mais simples, "A>100" e "B<50". Quando o Serviço de Integração de Dados divide um filtro, ele move os filtros simplificados para cima no pipeline do mapeamento, aproximando-os da origem. O Serviço de Integração de Dados move os filtros para cima no pipeline separadamente ao dividir o filtro.

O método de otimização de seleção antecipada é ativado por padrão quando você escolhe o nível normal ou completo do otimizador na Developer tool. O Serviço de Integração de Dados ignorará a otimização de seleção antecipada se uma transformação que aparece antes da transformação de Filtro tiver efeitos colaterais. O Serviço de Integração de Dados não pode determinar se a transformação de SQL, a transformação de Consumidor de Serviço da Web e a transformação de Java têm efeitos colaterais. Você poderá configurar a otimização de seleção antecipada para essas transformações se elas não apresentarem efeitos colaterais.

É possível desabilitar a seleção antecipada quando a otimização não melhora o desempenho. O Serviço de Integração de Dados ativa esse método de otimização por padrão.

Método de Otimização de Predicado Global

Quando o Serviço de Integração de Dados usa o método de otimização de predicado global, ele remove as linhas que podem ser eliminadas por filtragem o quanto antes no mapeamento. Isso reduz o número de linhas que precisam ser processadas pelo mapeamento. O método de otimização de predicado global inclui os métodos de otimização de predicado e de seleção antecipada.

Por exemplo, um mapeamento contém uma transformação de Associador com a condição de associação "A=B" e uma transformação de filtro com a condição de filtro "A>5". O Serviço de Integração de Dados pode adicionar "B>5" à condição de associação e aproximar a transformação de Filtro da origem.

O método de otimização de predicado global aplica expressões de predicado de maneira mais eficiente do que o método de otimização de predicado. O método de otimização de predicado global determina se ele pode simplificar ou gravar novamente as expressões para melhorar o desempenho. Ele também tenta aplicar as expressões de predicado o quanto antes no mapeamento para melhorar o desempenho dele.

O método de otimização de predicado global infere filtros e os aproxima da origem quando o mapeamento contém associadores aninhados ou ramificações com filtros em cada ramificação. Quando o Serviço de Integração de Dados usa o método de otimização de predicado global, ele separa os filtros, aproxima-os da origem ou remove os filtros em um mapeamento.

Método de Otimização de Ajuste de Ramificação

O Serviço de Integração de Dados pode aplicar o método de otimização de ajuste de ramificação a transformações que não contribuem com linhas para o destino no mapeamento.

O Serviço de Integração de Dados pode remover uma transformação de Filtro se a condição de filtro for avaliada como FALSE para as linhas de dados. Por exemplo, um mapeamento tem duas transformações de Filtro que filtram dados de duas origens relacionais. Uma transformação de Filtro tem a condição de filtro País=US e a outra transformação de Filtro tem a condição de filtro País=Canadá. Uma transformação de União une as duas origens relacionais e tem a condição de filtro País=US. O Serviço de Integração de Dados pode remover a transformação de Filtro com a condição de filtro País=Canadá do mapeamento.

A ferramenta Developer ativa o método de otimização de ajuste de ramificação por padrão quando você escolhe o nível do otimizador completo ou normal. Você poderá desativar o ajuste de ramificação se a otimização não melhorar o desempenho, configurando o nível do otimizador como mínimo ou nenhum.

Método de Otimização de Envio

Com a otimização de envio, o Serviço de Integração de Dados move a lógica de transformação de Filtro para a transformação imediatamente acima da transformação de Filtro no mapeamento. A otimização de envio melhora o desempenho ao reduzir o número de linhas que passam pelo mapeamento.

O Serviço de Integração de Dados não move a lógica de filtro para outra transformação quando essa transformação apresenta efeitos colaterais. O Serviço de Integração de Dados não pode determinar se a transformação SQL, a transformação de Consumidor de Serviço da Web e a transformação Java têm efeitos colaterais. Contudo, é possível configurar a transformação SQL, a transformação de Consumidor de Serviço da Web e a transformação Java para a otimização de envio.

Otimização de Empilhamento

Quando o Serviço de Integração de Dados aplica a otimização de empilhamento, ele envia a lógica de transformação ao banco de dados de origem. O Serviço de Integração de Dados converte a lógica de transformação em consultas SQL e envia essas consultas SQL ao banco de dados. O banco de dados de origem executa as consultas SQL para processar as transformações.

A otimização de empilhamento melhora o desempenho do mapeamento quando o banco de dados de origem pode processar a lógica de transformação com mais rapidez que o Serviço de Integração de Dados. O Serviço de Integração de Dados também lê menos dados da origem.

A quantidade de lógica de transformação que o Serviço de Integração de Dados envia ao banco de dados de origem depende do banco de dados, da lógica de transformação e da configuração de mapeamento. O Serviço de Integração de Dados processa toda lógica de transformação que ele não consegue enviar a um banco de dados.

Quando você aplica a otimização de empilhamento, o Serviço de Integração de Dados analisa o mapeamento otimizado da origem ao destino ou até atingir uma transformação descendente que ele não consegue enviar ao banco de dados de origem. O Serviço de Integração de Dados gera e executa uma consulta SELECT para cada origem com uma lógica de transformação empilhada. O Serviço de Integração de Dados também poderá gerar uma consulta INSERT se o destino tiver sido enviado para o banco de dados. O Serviço de Integração de Dados lê os resultados das consultas SQL e processa o restante das transformações no mapeamento.

O Serviço de Integração de Dados aplica a otimização de empilhamento a um mapeamento quando você seleciona o tipo de empilhamento nas propriedades de tempo de execução do mapeamento.

Você pode selecionar os seguintes tipos de empilhamento:

- Nenhum. Não selecione nenhum tipo de empilhamento para o mapeamento.
- Fonte. O Serviço de Integração de Dados tenta enviar o máximo possível da lógica de transformação para o banco de dados de origem.
- Completo. O Serviço de Integração de Dados envia a lógica de transformação completa para o banco de dados de origem.

Você também pode criar um parâmetro de cadeia para o tipo de empilhamento e usar os seguintes valores de parâmetro:

- Nenhum
- Origem
- Completo

Otimização de Empilhamento Total

Quando o Serviço de Integração de Dados aplica a otimização de empilhamento total, ele envia toda a lógica de transformação no mapeamento para o banco de dados de origem. Você pode configurar o empilhamento total nas propriedades de tempo de execução do mapeamento.

A otimização de empilhamento total é ideal quando a origem e o destino estão no mesmo banco de dados ou quando as transformações, como as transformações do Agregador e de Filtro, são processadas no banco de dados de origem e reduzem a quantidade de dados movidos. Por exemplo, se um mapeamento contiver uma origem e um destino da Teradata, configure a otimização de empilhamento total para enviar toda a lógica de transformação para processamento de um banco de dados de origem da Teradata para um banco de dados de destino da Teradata.

Ao configurar um mapeamento com uma transformação de Estratégia de Atualização para empilhamento total, você deve determinar a compatibilidade de empilhamento do mapeamento.

O Serviço de Integração de Dados pode empilhar um mapeamento com uma transformação de Estratégia de Atualização nas seguintes situações:

- Se a transformação de destino conectada à transformação de Estratégia de Atualização receber várias linhas que não têm a mesma chave.
- Se a transformação de destino conectada à transformação de Estratégia de Atualização receber várias linhas com a mesma chave que podem ser reordenadas.

O Serviço de Integração de Dados não pode empilhar um mapeamento com uma transformação de Estratégia de Atualização no seguinte cenário:

- Se a transformação de destino conectada à transformação de Estratégia de Atualização receber várias linhas com a mesma chave que não podem ser reordenadas.

Você também pode usar um parâmetro de compatibilidade de empilhamento no mapeamento. Você pode usar os seguintes valores de parâmetro:

- noMultipleRowsWithSameKeyOnTarget
- reorderAllowedForMultipleRowsWithSameKey
- reorderNotAllowedForRowsWithSameKey

O Serviço de Integração de Dados pode usar a otimização de empilhamento total nas seguintes origens:

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

Empilhamento de Origem

Quando o Serviço de Integração de Dados aplica o empilhamento de origem, ele analisa o mapeamento da origem ao destino ou até que ele atinja uma transformação downstream que não consegue enviar para o banco de dados de origem.

O Serviço de Integração de Dados gera e executa uma instrução SELECT com base na lógica de transformação para cada transformação que ele é capaz de enviar para o banco de dados. Em seguida, ele lê os resultados dessa consulta SQL e processa o restante das transformações.

Você poderá configurar um mapeamento para usar o empilhamento de origem se a origem e o destino residirem em bancos de dados diferentes. Por exemplo, se um mapeamento contiver uma origem da Teradata e um destino do Oracle, você poderá configurar o empilhamento de origem para enviar alguma lógica de transformação para processamento para a origem da Teradata.

Regras e Diretrizes para Otimização de Empilhamento

O Serviço de Integração de Dados pode enviar a lógica de transformação para o banco de dados de origem.

As seguintes regras e diretrizes se aplicam à otimização de empilhamento:

- O Serviço de Integração de Dados poderá enviar a lógica de transformação de Pesquisa e de Associador para o banco de dados de origem se as origens estiverem no mesmo sistema de gerenciamento de banco de dados e usarem conexões idênticas.
- O Serviço de Integração de Dados não pode enviar a lógica de transformação para uma origem que tenha um tipo de dados binário.
- O Serviço de Integração de Dados desativa a otimização de empilhamento quando existe uma fonte de dados IBM DB2 e a precisão de colunas é de 28 a 31 dígitos para o tipo de dados Decimal.
- Por padrão, o Serviço de Integração de Dados ativa a otimização de empilhamento para um Serviço de Dados SQL ou um Serviço da Web. Você não pode desativar a otimização de empilhamento para um Serviço de Dados SQL ou um Serviço da Web.
- O Serviço de Integração de Dados não pode enviar uma transformação de Agregador que contenha uma expressão com funções agregadas e não agregadas em uma porta que não seja de agrupamento.

Leitura de Passagem Única

A leitura de passagem única permite que você preencha vários destinos com um objeto de dados personalizado. Considere o uso da leitura de passagem única quando existem várias sessões que usam as mesmas origens.

Considere as seguintes soluções para afunilamentos de leitura de passagem única:

Combinar a lógica de transformação para cada mapeamento em um único mapeamento e usar um objeto de dados personalizado para cada origem.

O Data Integration Service lê cada fonte uma vez e depois envia os dados em pipelines separados. Uma linha específica pode ser usada por todos os pipelines, por qualquer combinação de pipelines ou por nenhum pipeline.

Por exemplo, você tem a tabela de origem Compras e usa essa origem diariamente para realizar uma agregação e uma classificação. Se você inserir as transformações de Agregador e de Classificação em mapeamentos separados, forçará o Data Integration Service a ler a mesma tabela de origem duas vezes. No entanto, se você incluir a lógica de agregação e de classificação em um mapeamento com um

qualificador de origem, o Data Integration Service lerá a tabela de origem Compras uma vez e enviará os dados adequados aos pipelines separados.

Fatorar funções comuns de mapeamentos.

Quando você altera mapeamentos para tirar proveito da leitura de passagem única, é possível otimizar esse recurso fatorando funções comuns de mapeamentos. Por exemplo, se você precisar subtrair uma porcentagem das portas de Preço tanto para as transformações de Agregador quanto para as transformações de Classificação, será possível minimizar o trabalho subtraindo a porcentagem antes de dividir o pipeline. Você pode usar uma transformação de Expressão para subtrair a porcentagem e depois dividir o mapeamento após a transformação.

Otimização de Filtros

Você pode otimizar mapeamentos filtrando dentro de um objeto de dados personalizado e inserindo filtros com antecedência no mapeamento.

Considere as seguintes soluções para afunilamentos de filtros:

Usar um filtro em um objeto de dados personalizado para remover as linhas na origem.

Se você filtrar as linhas do mapeamento, poderá melhorar a eficiência filtrando com antecedência no fluxo de dados. Usar um filtro em um objeto de dados personalizado para remover as linhas na origem. O objeto de dados personalizado limita o conjunto de linhas extraído de uma origem relacional.

Se você não puder usar um filtro no objeto de dados personalizado, use uma transformação de Filtro e mova-a o mais próximo possível do objeto de dados personalizado para remover dados desnecessários com antecedência no fluxo de dados. A transformação de Filtro limita o conjunto de linhas enviado a um destino.

Use um filtro em uma transformação de Estratégia de Atualização se você não precisar manter linhas rejeitadas.

Para melhorar o desempenho do mapeamento, você também pode usar uma transformação de Filtro para descartar linhas rejeitadas de uma transformação de Estratégia de Atualização se não precisar manter as linhas rejeitadas.

Evitar expressões complexas em condições de filtro.

Evite usar expressões complexas em condições de filtro. Para otimizar transformações de Filtro, use expressões simples de inteiro ou de verdadeiro/falso na condição do filtro.

A transformação de Filtro filtra dados dentro de um mapeamento. A transformação de Filtro filtra linhas de qualquer tipo de origem. O objeto de dados personalizado filtra linhas de origens relacionais. A transformação de Filtro filtra linhas de qualquer tipo de origem.

Otimização de Conversão de Tipos de Dados

Você pode melhorar o desempenho eliminando conversões de tipos de dados desnecessárias. Por exemplo, se um mapeamento mover dados de uma coluna Inteiro para uma coluna Decimal e retornar para uma coluna Inteiro, a conversão de tipos de dados desnecessária reduzirá o desempenho. Quando possível, elimine conversões de tipos de dados desnecessárias dos mapeamentos.

Considere as seguintes soluções para afunilamentos de conversão de tipos de dados:

Usar valores inteiros no lugar de outros tipos de dados ao realizar comparações usando transformações de Pesquisa e de Filtro.

Por exemplo, muitos bancos de dados armazenam informações de CEP dos Estados Unidos como um tipo de dados Char ou Varchar. Se você converter os dados de CEP em um tipo de dados de Inteiro, o banco de dados de pesquisa armazenará o CEP 94303-1234 como 943031234. Isso ajuda a aumentar a velocidade das comparações de pesquisa com base no CEP.

Converter as datas de origem em strings por meio de conversões de porta a porta para aumentar o desempenho do mapeamento.

Você pode deixar as portas em destinos como strings ou alterá-las para portas de Data/Hora.

Rastreamento de Erros

Para melhorar o desempenho, reduza o número de eventos de log gerados pelo Data Integration Service quando ele executar o mapeamento. Melhore o desempenho do mapeamento atualizando o nível do otimizador de mapeamento por meio da configuração de mapeamento ou de propriedades de implantação de mapeamento. Use o método de otimização baseada em custos para otimizar mapeamentos.

Considere as seguintes soluções para afunilamentos de rastreamento de erros:

Definir o nível de rastreamento nas propriedades de mapeamento como Conciso

Se um mapeamento contiver muitos erros de transformação, e você não precisar corrigi-los, defina o nível de rastreamento nas propriedades de mapeamento para Conciso. Nesse nível de rastreamento, o Data Integration Service não grava mensagens de erro ou informações em nível de linha para dados rejeitados.

Se você precisar depurar o mapeamento e definir o nível de rastreamento como Detalhado, poderá haver um impacto negativo significativo no desempenho quando o mapeamento for executado. Não use o rastreamento Detalhado quando ajustar o desempenho. O nível de rastreamento do mapeamento substitui qualquer outro nível de rastreamento específico de transformação contido no mapeamento. Isso não é recomendado como uma resposta a longo prazo para altos níveis de erros de transformação.

Alterar o nível do otimizador para o mapeamento.

Se um mapeamento demorar muito para ser executado, convém alterar o nível do otimizador para esse mapeamento. O nível do otimizador determina quais métodos de otimização o Data Integration Service aplica ao mapeamento em tempo de execução.

Você define o nível do otimizador para um mapeamento na configuração do mapeamento ou nas propriedades de implantação do mapeamento. O Data Integration Service aplica diferentes níveis de otimizador ao mapeamento, dependendo de como você executa esse mapeamento.

Usar o método de otimização baseada em custos.

O método de otimização baseada em custos faz com que o Data Integration Service avalie um mapeamento, gere mapeamentos semanticamente equivalentes e execute o mapeamento com o melhor desempenho. Esse método é mais eficaz para mapeamentos que contêm várias transformações de Associador. Ele reduz o tempo de execução para mapeamentos que realizam operações de associação interna adjacentes e não classificadas.

Mapeamentos semanticamente equivalentes são mapeamentos que realizam funções idênticas e produzem os mesmos resultados. Para gerar mapeamentos semanticamente equivalentes, o Data Integration Service divide o mapeamento original em fragmentos. O Data Integration Service determina então quais fragmentos de mapeamento ele pode otimizar.

CAPÍTULO 6

Otimização de Mapeamento Particionado

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Otimização de Mapeamento Particionado, 54](#)
- [Usar Várias CPUs, 55](#)
- [Aumentar o Valor Máximo de Paralelismo, 55](#)
- [Otimizar Arquivos Simples para Particionamento, 56](#)
- [Otimizar Bancos de Dados Relacionais para Particionamento, 57](#)
- [Otimizar Transformações para Particionamento, 58](#)

Visão Geral da Otimização de Mapeamento Particionado

Se você tiver a opção de particionamento, poderá ativar o Serviço de Integração de Dados para maximizar o paralelismo quando ele executar mapeamentos. Quando você maximiza o paralelismo, o Serviço de Integração de Dados divide dinamicamente os dados subjacentes em partições e processa cada partição simultaneamente.

Se os mapeamentos processarem grandes conjuntos de dados ou tiverem transformações que executam cálculos complicados, eles poderão levar um longo tempo para serem processados e poderão resultar em baixa taxa de transferência de dados. Quando você ativa o particionamento nesses mapeamentos, o Serviço de Integração de Dados usa segmentos adicionais para processar o mapeamento.

Você pode otimizar o desempenho dos mapeamentos particionados realizando as seguintes tarefas:

- Use várias CPUs nos nós que executam mapeamentos.
- Aumente o valor de paralelismo máximo do Serviço de Integração de Dados.
- Configure propriedades nos objetos de dados de arquivo simples.
- Configure bancos de dados relacionais para otimizar o particionamento.
- Configure propriedades nas transformações.

Usar Várias CPUs

Aumentar o número de segmentos de processamento aumenta a carga nos nós que executam mapeamentos. Se os nós contiverem largura de banda de CPU suficiente, o processamento simultâneo de linhas de dados em um mapeamento pode otimizar o desempenho de mapeamento.

O Serviço de Integração de Dados pode usar várias CPUs para processar um mapeamento que contenha várias partições. O número de CPUs que o serviço usa depende de fatores tais como o número de pontos de partição, o número de segmentos criados para cada fase de pipeline e a quantidade de recursos necessários para processar o mapeamento. Um mapeamento simples é executado mais rapidamente em duas partições, mas normalmente exige duas vezes mais CPU do que quando o mapeamento é executado em uma única partição.

Aumentar o Valor Máximo de Paralelismo

O paralelismo máximo determina o número máximo de segmentos paralelos que podem processar uma única fase de pipeline. Configurar a propriedade **Máximo de Paralelismo** do Serviço de Integração de Dados com base nos recursos de hardware disponíveis. Quando você aumenta o valor de paralelismo máximo, pode reduzir o tempo de processamento.

Considere as seguintes diretrizes quando você configurar o valor de paralelismo máximo:

Aumente o valor com base no número de CPUs disponíveis.

Aumenta o valor de paralelismo máximo com base no número de CPUs disponíveis nos nós onde há execução de mapeamentos. Quando você aumenta o valor do paralelismo máximo, o Serviço de Integração de Dados usa mais segmentos para executar o mapeamento e fornece mais CPUs. Um mapeamento simples é executado com mais rapidez em duas partições, mas normalmente requer duas vezes mais a quantidade de CPU que quando o mapeamento é executado em uma única partição.

Considere o número total de segmentos de processamento.

Considere o número total de segmentos de processamento ao definir o valor de paralelismo máximo. Se um mapeamento complexo resultar em vários pontos de partição adicionais, o Serviço de Integração de Dados poderá usar mais segmentos de processamento do que a CPU é capaz de processar.

O número total de segmentos de processamento é igual ao valor de paralelismo máximo.

Considere os outros trabalhos que o Serviço de Integração de Dados deve executar.

Se você configurar o paralelismo máximo de forma que cada mapeamento usar um grande número de segmentos, menos segmentos estarão disponíveis para o Serviço de Integração de Dados executar tarefas adicionais.

Opcionalmente, altere o valor de um mapeamento.

Por padrão, o paralelismo máximo para cada mapeamento é definido como Automático. Cada mapeamento usa o valor de paralelismo máximo definido para o Serviço de Integração de Dados.

Na Developer tool, os desenvolvedores podem alterar o valor de paralelismo máximo nas propriedades de tempo de execução de mapeamento para definir um valor máximo para um determinado mapeamento. Quando o paralelismo máximo é definido como valores inteiros diferentes no Serviço de Integração de Dados e no mapeamento, o Serviço de Integração de Dados usa o valor mínimo dos dois.

Otimizar Arquivos Simples para Particionamento

Quando um mapeamento que está ativado para particionamento lê a partir de uma origem de arquivo simples ou grava em um destino de arquivo simples, o Serviço de Integração de Dados pode usar vários segmentos para ler ou gravar nesse arquivo.

Otimizar Origens de Arquivo Simples para Particionamento

Para obter o desempenho ideal durante o uso de vários segmentos para ler um arquivo simples, configure o objeto de dados de arquivo simples para otimizar a produtividade, em vez de preservar a ordem das linhas.

Considere a seguinte solução para reduzir gargalos de origens de arquivo simples particionadas:

Configurar o particionamento simultâneo de leitura do objeto de dados de arquivo simples para otimizar a produtividade.

Nas propriedades avançadas do objeto de dados de arquivo simples, defina a propriedade **Particionamento Simultâneo de Leitura** para otimizar a produtividade. Quando você otimiza a produtividade, o Serviço de Integração de Dados não preserva a ordem das linhas porque ele não lê as linhas no arquivo ou na lista de arquivos de forma sequencial.

Otimizar Destinos de Arquivo Simples para Particionamento

Para obter o desempenho ideal ao usar vários segmentos para gravar em um arquivo simples, configure as partições para gravar a saída de destino em arquivos separados e configurar vários diretórios de destino.

Considere as seguintes soluções para reduzir gargalos de destinos de arquivo simples particionados:

Configurar as partições para gravar a saída de destino em arquivos separados.

Nas propriedades avançadas do objeto de dados de arquivo simples, defina a propriedade **Tipo de Mesclagem** como **Não mesclar**. O Serviço de Integração de Dados grava simultaneamente a saída de destino em um arquivo separado para cada partição. Se você exigir dados de destino mesclados, o tipo de mesclagem simultânea otimiza o desempenho mais do que o tipo de mesclagem sequencial.

Configure vários diretórios de destino.

Quando vários segmentos gravam em um único diretório, o mapeamento pode encontrar um gargalo devido à contenção de entrada/saída (E/S). Uma contenção de E/S pode ocorrer quando os segmentos gravam dados no sistema de arquivos ao mesmo tempo. Quando você configura vários diretórios, o Serviço de Integração de Dados determina o diretório de saída de cada segmento ao modo round-robin.

Configure os diretórios de arquivo de saída nas propriedades avançadas do objeto de dados de arquivo simples. Use o valor do parâmetro do sistema padrão TargetDir se um administrador tiver inserido vários diretórios separados por ponto e vírgula na propriedade **Diretório de Destino** do Serviço de Integração de Dados na ferramenta Administrator. Se preferir, você pode inserir um valor diferente para configurar vários diretórios de arquivo de saída específicos do objeto de dados de arquivo simples.

Otimizar Bancos de Dados Relacionais para Particionamento

Quando um mapeamento que está ativado para particionamento lê ou grava em um banco de dados relacional IBM DB2 for LUW ou Oracle, o Serviço de Integração de Dados pode usar vários segmentos para ler ou gravar no destino relacional.

Para otimizar o desempenho durante o uso de vários segmentos para ler ou gravar em um banco de dados relacional do DB2 para LUW ou do Oracle, você pode particionar as tabelas de origem e de destino.

Nota: Se um mapeamento ler ou gravar em um banco de dados relacional diferente do DB2 for LUW ou Oracle, o Serviço de Integração de Dados usará um segmento leitor ou um segmento gravador.

Otimizar o Banco de Dados de Origem para Particionamento

Para obter o desempenho ideal durante o uso de vários segmentos para ler um banco de dados de origem do DB2 para LUW ou do Oracle, verifique se a tabela de origem está particionada e configurada para aceitar consultas paralelas.

Para otimizar o banco de dados de origem para particionamento, execute as seguintes tarefas:

Adicione partições de banco de dados à origem.

Adicione partições de banco de dados à origem relacional para aumentar a velocidade da consulta do Serviço de Integração de Dados que lê a origem. Se a origem não tiver partições de banco de dados, o Serviço de Integração de Dados usará um thread para ler da origem.

Permita consultas paralelas.

Os bancos de dados relacionais podem ter opções que permitem consultas paralelas no banco de dados. Consulte a documentação do banco de dados para essas opções. Se essas opções não estiverem ativadas, o Serviço de Integração de Dados executará várias instruções SELECT de partição serialmente.

Separe os dados em diferentes espaços de tabela.

Cada banco de dados fornece uma opção para separar os dados em diferentes espaços de tabela. Cada espaço de tabela pode fazer referência a um sistema de arquivos exclusivo, que evita qualquer contenção de E/S entre as partições.

Aumente o número máximo de sessões permitido para o banco de dados.

O Serviço de Integração de Dados cria uma conexão separada para o banco de dados de origem de cada partição. Aumente o número máximo de sessões de modo que o banco de dados possa conter um número maior de conexões simultâneas.

Otimizar o Banco de Dados de Destino para Particionamento

Para obter o desempenho ideal durante o uso de vários segmentos para gravar em um banco de dados de destino do DB2 para LUW ou do Oracle, verifique se a tabela de destino está configurada para inserir linhas em paralelo.

Para otimizar o banco de dados de destino para particionamento, execute as seguintes tarefas:

Adicione partições de banco de dados a um destino do DB2 para LUW.

O Serviço de Integração de Dados pode usar vários threads para gravar em um destino do DB2 para LUW que não tem partições de banco de dados. No entanto, você poderá otimizar o desempenho de carregamento quando o destino tiver partições de banco de dados. Nesse caso, cada thread de gravador

se conecta ao nó do DB2 para LUW que contém a partição de banco de dados. Como os segmentos de gravador se conectam a um nó do DB2 para LUW diferente em vez de todos os segmentos se conectarem ao único nó mestre, o desempenho aumenta.

Permita inserções paralelas.

Os bancos de dados relacionais podem ter opções que permitem inserções paralelas no banco de dados. Consulte a documentação do banco de dados para essas opções. Por exemplo, defina a opção `db_writer_processes` em um banco de dados Oracle e a opção `max_agents` em um banco de dados do DB2 para LUW para permitir inserções paralelas.

Separe os dados em diferentes espaços de tabela.

Cada banco de dados fornece uma opção para separar os dados em diferentes espaços de tabela. Cada espaço de tabela pode fazer referência a um sistema de arquivos exclusivo, que evita qualquer contenção de E/S entre as partições.

Aumente o número máximo de sessões permitido para o banco de dados.

O Serviço de Integração de Dados cria uma conexão separada para o banco de dados de destino de cada partição. Aumente o número máximo de sessões de modo que o banco de dados possa conter um número maior de conexões simultâneas.

Defina as opções para aprimorar a escalabilidade do banco de dados.

Bancos de dados relacionais podem ter opções que aprimoram a escalabilidade. Por exemplo, desative o log de arquivo e as estatísticas limitadas em um banco de dados Oracle para aprimorar a escalabilidade.

Otimizar Transformações para Particionamento

Quando o Serviço de Integração de Dados usa vários segmentos para executar uma transformação do Agregador, do Associador, de Classificação ou do Classificador, o serviço usa o particionamento do cache para dividir o tamanho do cache nos segmentos. Para otimizar o desempenho do particionamento do cache, configure vários diretórios de cache.

Nota: Uma transformação de Pesquisa pode usar somente um único diretório de cache.

Considere a seguinte solução para reduzir os gargalos das transformações do Agregador, do Associador, de Classificação e do Classificador particionadas:

Configure vários diretórios de cache.

O particionamento do cache cria um cache separado para cada partição que processa uma transformação do Agregador, do Associador, de Classificação ou de Classificador. Durante o particionamento do cache, cada partição armazena dados diferentes em um cache separado. Cada cache contém as linhas necessárias nessa partição. O particionamento do cache otimiza o desempenho do mapeamento porque cada segmento consulta um cache separado em paralelo.

Se o tamanho do cache for menor que a quantidade de memória necessária para executar a transformação, os segmentos de transformação gravarão no diretório de cache para armazenar os valores de overflow nos arquivos de cache. Quando vários segmentos gravam em um único diretório, o mapeamento pode encontrar um afunilamento devido à contenção de E/S. Uma contenção de E/S pode ocorrer quando os segmentos gravam dados no sistema de arquivos ao mesmo tempo. Quando você configura vários diretórios de cache, o Serviço de Integração de Dados determina o diretório de cache de cada segmento de transformação ao modo round-robin.

Em uma transformação do Agregador, do Associador ou de Classificação, configure os diretórios de cache na propriedade avançada **Diretório de Cache**. Use o valor do parâmetro do sistema padrão CacheDir se um administrador tiver inserido vários diretórios separados por ponto e vírgula na propriedade **Diretório de Cache** do Serviço de Integração de Dados na ferramenta Administrator. Se preferir, você pode inserir um valor diferente para configurar vários diretórios de cache específicos para a transformação.

Em uma transformação do Classificador, configure o diretórios de cache na propriedade avançada **Diretório de Trabalho**. Use o valor do parâmetro do sistema padrão TempDir se um administrador tiver inserido vários diretórios separados por ponto e vírgula na propriedade **Diretórios Temporários** do Serviço de Integração de Dados na ferramenta Administrator. Se preferir, você pode inserir um valor diferente para configurar vários diretórios de cache específicos para a transformação.

CAPÍTULO 7

Otimização do Tempo de Execução

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Otimização do Tempo de Execução, 60](#)
- [Otimização de Serviços de Aplicativo, 60](#)
- [Estatísticas de Monitoramento, 63](#)
- [Alocação de Memória, 64](#)
- [Armazenamento em Cache do Objeto de Dados, 65](#)
- [Otimização do Sistema, 68](#)

Visão Geral da Otimização do Tempo de Execução

Ative os recursos de desempenho e ajuste as propriedades do Serviço de Integração de Dados para otimizar o desempenho dos mapeamentos.

Use as seguintes técnicas de otimização na ferramenta Administrator para obter os melhores resultados de desempenho com base nos seus requisitos:

- Otimizar processos de serviços de aplicativo.
- Configurar estatísticas de monitoramento para monitorar afunilamentos do sistema.
- Alocar memória para obter o desempenho ideal do sistema.
- Configurar o armazenamento em cache de objetos de dados.
- Otimizar o sistema para evitar atrasos do sistema e reduzir a velocidade de acesso ao disco.

Otimização de Serviços de Aplicativo

Otimize o processo de serviços de aplicativo quando o desempenho é afetado. Você pode otimizar o Analyst Service, o Data Integration Service e o Model Repository Service.

Otimização do Serviço Analyst

Ajuste o Serviço Analyst para otimizar o desempenho. Você pode configurar a propriedade do processo do Serviço Analyst para memória, minimizar a latência da rede e definir as configurações de carregamento de arquivos simples da ferramenta Analyst para melhorar o desempenho do serviço.

Considere as seguintes soluções para afunilamentos do Serviço Analyst:

Configurar a ferramenta Analyst de forma a se conectar a uma localização de caminho de rede para carregar arquivos simples com mais de 10 MB.

O desempenho do processo do Serviço Analyst pode diminuir quando os analistas carregam arquivos simples com mais de 10 MB para o diretório de instalação da Informatica na máquina na qual a ferramenta Analyst é executada. Isso pode afetar o espaço em disco e o desempenho da rede.

Carregar arquivos simples menores que 10 MB no diretório de Instalação Informatica a partir da ferramenta Analyst.

O desempenho do processo do Serviço Analyst pode diminuir quando os analistas carregam arquivos simples com mais de 10 MB no diretório de Instalação Informatica a partir da ferramenta Analyst. Isso pode afetar tanto o espaço em disco quanto o desempenho da rede.

Aumentar a propriedade Tamanho Máximo do Heap para o processo do Serviço Analyst.

O processo do Serviço Analyst pode consumir grandes quantidades de memória durante o processamento de um grande número de usuários simultaneamente conectados. Isso pode resultar na abertura de um grande número de conexões de rede entre o Serviço Analyst e outros serviços, como o Serviço de Integração de Dados ou o Serviço de Repositório do Modelo.

Use a ferramenta Administrator para configurar a propriedade Tamanho Máximo do Heap como um valor maior nas Propriedades Avançadas do processo do Serviço Analyst.

Exportar especificações de mapeamento grandes para uma tabela ou exportar para um arquivo simples e truncar esse arquivo.

O processo do Serviço Analyst pode ter impacto sobre o desempenho quando os analistas exportam especificações de mapeamento grandes como arquivos simples a partir da ferramenta Analyst.

Otimização do Serviço de Integração de Dados

Ajuste o processo do Serviço de Integração de Dados para melhorar o desempenho do serviço. Você pode configurar as propriedades do processo do Serviço de Integração de Dados para memória. Você pode configurar cada serviço da Web e serviço de dados SQL que é executado em um Serviço de Integração de Dados para lidar com solicitações simultâneas.

Considere as seguintes soluções para afunilamentos do Serviço de Integração de Dados:

Configurar a propriedade Tamanho Máximo do Heap para o processo do Serviço de Integração de Dados.

O Serviço de Integração de Dados pode consumir grandes quantidades de memória durante o processamento de serviços de dados SQL e serviços da Web.

Use a ferramenta Administrator para configurar a propriedade Tamanho Máximo do Heap como um valor maior nas Propriedades Avançadas do processo do Serviço de Integração de Dados.

Configurar a propriedade Tempo de Atividade do DTM do serviço da Web para o Serviço de Integração de Dados.

O Serviço de Integração de Dados consome recursos do sistema para gerar um processo do DTM para cada solicitação de serviço da Web. Configure o Serviço de Integração de Dados para usar uma instância do DTM de forma a processar mais de uma solicitação de serviço da Web.

Use a ferramenta Administrator para configurar a propriedade Tempo de Atividade do DTM do serviço da Web para o Serviço de Integração de Dados.

Configurar as opções de execução nas propriedades do Serviço de Integração de Dados e nas propriedades de serviços da Web e serviços de dados SQL para solicitações simultâneas.

O Serviço de Integração de Dados, cada serviço de dados SQL e cada serviço da Web que é executado no Serviço de Integração de Dados consomem recursos de memória e do sistema para cada solicitação simultânea.

Para configurar o número de solicitações simultâneas que podem ser aceitas pelo Serviço de Integração de Dados, por cada serviço de dados SQL e por cada serviço da Web, configure as propriedades do Serviço de Integração de Dados e as propriedades do serviço da Web.

Use a ferramenta Administrator para configurar as seguintes opções e propriedades para o Serviço de Integração de Dados, o serviço da Web e o serviço de dados SQL:

- Configurar as opções de execução para o Serviço de Integração de Dados.
- Configurar a propriedade Nº Máximo de Conexões Simultâneas para cada serviço de dados SQL nas propriedades SQL do processo do Serviço de Integração de Dados.
- Configurar as propriedades Número Máximo de Solicitações de Registros Acumulados e Número Máximo de Solicitações Simultâneas para cada serviço da Web nas propriedades de configuração HTTP do processo do Serviço de Integração de Dados.

Desativar o nível de rastreamento do serviço da Web.

O número de arquivos de log do serviço da Web que são gravados e mantidos pelo Serviço de Integração de Dados pode diminuir o desempenho.

Use a ferramenta Administrator para configurar o nível de rastreamento do serviço da Web de forma a reduzir a quantidade de arquivos de log de tempo de execução do serviço da Web que o Serviço de Integração de Dados armazena em disco.

Otimização do Serviço de Repositório do Modelo

Ajuste o Serviço de Repositório do Modelo para melhorar o desempenho. Você pode configurar a propriedade do processo do Serviço de Repositório do Modelo para memória e para minimizar a latência da rede.

Considere as seguintes soluções para afunilamentos do Serviço de Repositório do Modelo:

Hospedar o banco de dados do repositório do Modelo na mesma máquina do Serviço de Repositório do Modelo.

O desempenho do processo do Serviço de Repositório do Modelo pode ser afetado quando o banco de dados do repositório do Modelo está hospedado em um servidor remoto. Operações do Serviço de Repositório do Modelo que exigem comunicação entre o repositório do Modelo e o Serviço de Repositório do Modelo em uma rede de alta latência podem diminuir o desempenho do Serviço de Repositório do Modelo.

Aumentar a propriedade Tamanho Máximo do Heap para o processo do Serviço de Repositório do Modelo.

O processo do Serviço de Repositório do Modelo pode consumir grandes quantidades de memória durante o processamento de um grande número de usuários simultaneamente conectados. Isso pode resultar na abertura de um grande número de conexões de rede entre o Serviço de Repositório do Modelo e outros serviços, como o Serviço de Integração de Dados ou o Serviço Analyst.

Use a ferramenta Administrator para configurar a propriedade Tamanho Máximo do Heap como um valor maior nas Propriedades Avançadas do processo do Serviço de Repositório do Modelo.

Estatísticas de Monitoramento

O monitoramento é uma função de domínio realizada pelo Gerenciador de Serviços. O Gerenciador de Serviços armazena a configuração de monitoramento no repositório do Modelo. Use a guia **Monitorar** na ferramenta **Administrador** para monitorar afunilamentos do sistema, como o número total de trabalhos em execução, com falha, cancelados e concluídos que são executados em um serviço selecionado.

Considere a seguinte solução para afunilamentos de estatísticas de monitoramento:

Configurar o domínio para definir o monitoramento.

Quando você define o monitoramento, o Serviço de Integração de Dados armazena estatísticas persistentes e relatórios de monitoramento no repositório do Modelo. Estatísticas persistentes são informações históricas sobre objetos de integração que foram executados anteriormente. Os relatórios de monitoramento mostram métricas-chave sobre um objeto de integração.

Definir as configurações de monitoramento do domínio para especificar o repositório do Modelo que armazena as estatísticas em tempo de execução sobre os objetos implantados no Serviço de Integração de Dados. As configurações de monitoramento se aplicam a todos os Serviços de Integração de Dados no domínio e podem afetar o desempenho do serviço.

A seguinte tabela descreve as configurações de monitoramento que podem afetar o desempenho dos serviços:

Opção	Descrição
Preservar dados históricos de resumo	Número de dias que o repositório do modelo salva dados de média. Se a limpeza for desativada, o repositório do modelo salvará os dados indefinidamente. O padrão é 180. O mínimo é 0. O máximo é 366.
Preservar dados históricos detalhados	Número de dias que o repositório do modelo salva dados por minuto. Se a limpeza for desativada, o repositório do modelo salvará os dados indefinidamente. O padrão é 14. O mínimo é 1. O máximo é 14.
Limpar Estatísticas a Cada	Intervalo, em dias, em que o Serviço de Repositório do Modelo limpa dados mais antigos que os valores configurados na opção Preservar Dados Históricos . O padrão é 1 dia.
Dias às	Hora do dia em que o Serviço de Repositório do Modelo limpa estatísticas. O padrão é 1:00 a.m.
Número Máximo de Registros Classificáveis	Número máximo de registros que podem ser classificados na guia Monitorar . Se o número de registros na guia Monitorar for maior do que este valor, você somente poderá classificar por Hora de Início e Hora de Término . O padrão é 3.000.
Atraso Máximo para Notificações de Atualização	Período máximo de tempo, em segundos, que o Serviço de Integração de Dados coloca em buffers as estatísticas antes de armazená-las no repositório do modelo e exibi-las na guia Monitorar . Se o Serviço de Integração de Dados for desativado inesperadamente antes de armazenar as estatísticas no repositório do modelo, as estatísticas serão perdidas. O padrão é 10.
Mostrar Milissegundos	Inclua milissegundos nos campos de data e hora na guia Monitorar .

Alocação de Memória

Para otimizar o desempenho do mapeamento, configure as propriedades da memória do Serviço de Integração de Dados na ferramenta Administrator.

A seguinte tabela descreve a propriedade memória máxima por solicitação do Módulo de Serviço de Mapeamento:

Propriedade	Descrição
Memória Máxima por Solicitação	<p>O comportamento da Memória Máxima por Solicitação depende das seguintes configurações do Serviço de Integração de Dados:</p> <ul style="list-style-type: none">- O serviço executa tarefas em processos locais ou remotos separados ou o tamanho máximo de memória da propriedade do serviço é 0 (padrão). <p>A Memória Máxima por Solicitação é a quantidade máxima de memória, em bytes, que o Serviço de Integração de Dados pode alocar em todas as transformações que usam o modo de cache automático em uma única solicitação. O serviço aloca memória separadamente nas transformações com um tamanho do cache específico. O total de memória usada pela solicitação pode exceder o valor de Máximo de Memória por Solicitação.</p> <ul style="list-style-type: none">- O serviço executa tarefas no processo do Serviço de Integração de Dados e o tamanho máximo de memória da propriedade do serviço é 0 (padrão). <p>A Memória Máxima por Solicitação é a quantidade máxima de memória, em bytes, que o Serviço de Integração de Dados pode alocar em uma única solicitação. O total de memória usada pela solicitação não pode exceder o valor de Máximo de Memória por Solicitação.</p> <p>O padrão é 536.870.912.</p>

A seguinte tabela descreve as opções de execução do Serviço de Integração de Dados:

Propriedade	Descrição
Tamanho Máximo do Pool de Execução	<p>Número máximo de trabalhos que cada processo do Serviço de Integração de Dados pode executar simultaneamente. Os trabalhos incluem visualizações de dados, mapeamentos, trabalhos de criação de perfil, consultas SQL e solicitações de serviço da Web. Por exemplo, uma grade do Serviço de Integração de Dados inclui três processos de serviço em execução. Se você definir o valor como 10, cada processo do Serviço de Integração de Dados poderá executar até 10 trabalhos simultaneamente. Um total de 30 trabalhos pode ser executado simultaneamente na grade. O padrão é 10.</p> <p>Nota: Quando você aumenta o valor do tamanho do pool, o Serviço de Integração de Dados usa mais recursos de hardware, como CPU, memória e E/S do sistema. Defina esse valor com base nos recursos disponíveis na máquina. Por exemplo, considere o número de CPUs na máquina que hospeda o Serviço de Integração de Dados e a quantidade de memória que está disponível para o Serviço de Integração de Dados.</p>
Tamanho Máximo de Memória	<p>Quantidade máxima de memória, em bytes, que o Serviço de Integração de Dados pode alocar para a execução de todas as solicitações simultaneamente quando o serviço executa tarefas em um processo do Serviço de Integração de Dados. Quando o Serviço de Integração de Dados executa tarefas em processos locais ou remotos separados, o serviço ignora esse valor. Se você não quiser limitar a quantidade de memória que o Serviço de Integração de Dados pode alocar, defina essa propriedade como 0.</p> <p>Se o valor for maior que 0, o Serviço de Integração de Dados usará a propriedade para calcular o máximo de memória total permitido para a execução de todas as solicitações simultaneamente. O Serviço de Integração de Dados calcula o máximo de memória total da seguinte forma:</p> <p>Tamanho Máximo de Memória + Tamanho Máximo do Heap + memória necessária para carregar componentes do programa</p> <p>O padrão é 0.</p> <p>Nota: Se você executar perfis ou mapeamentos de qualidade de dados, defina essa propriedade como 0.</p>

A seguinte tabela descreve a propriedade Tamanho Máximo do Heap para o processo do Serviço de Integração de Dados:

Propriedade	Descrição
Tamanho Máximo do Heap	<p>Quantidade de RAM alocada na Máquina Virtual Java (JVM) que executa o Serviço de Integração de Dados. Use essa propriedade para melhorar o desempenho. Acrescente uma das seguintes letras ao valor para especificar as unidades:</p> <ul style="list-style-type: none"> - b para bytes. - k para kilobytes. - m para megabytes. - g para gigabytes. <p>O padrão é 640 megabytes.</p> <p>Nota: Considere aumentar o tamanho do heap quando o Serviço de Integração de Dados necessitar processar grandes quantidades de dados.</p>

Armazenamento em Cache do Objeto de Dados

O Serviço de Integração de Dados usa o cache do objeto de dados para acessar objetos de dados lógicos pré-criados e tabelas virtuais. Ative o cache do objeto de dados para melhorar o desempenho de

mapeamentos, as consultas de serviço de dados SQL e as solicitações de serviço da Web que incluem objetos de dados lógicos e tabelas virtuais.

Por padrão, o Serviço de Integração de Dados extrai dados de origem e cria objetos de dados obrigatórios quando ele executa um mapeamento, uma consulta de serviço de dados SQL ou uma solicitação de serviço da Web. Quando você ativa o cache do objeto de dados, o Serviço de Integração de Dados pode usar objetos de dados lógicos e tabelas virtuais.

Realize as seguintes etapas para configurar o cache do objeto de dados para objetos de dados lógicos e tabelas virtuais em um aplicativo:

1. Configure a conexão de banco de dados do cache do objeto de dados nas propriedades do cache do Serviço de Integração de Dados.
2. Ative o armazenamento em cache nas propriedades de objetos de dados lógicos ou tabelas virtuais em um aplicativo.

Por padrão, o componente do Gerenciador de Cache do Objeto de Dados do Serviço de Integração de Dados gerencia as tabelas de cache dos objetos de dados lógicos e das tabelas virtuais no banco de dados do cache do objeto de dados. Quando o Gerenciador de Cache do Objeto de Dados gerencia o cache, ele insere todos os dados nas tabelas de cache com cada atualização. Se você desejar uma atualização incremental das tabelas de cache, poderá escolher gerenciar as tabelas de cache sozinho usando um cliente de banco de dados ou outra ferramenta externa. Depois de ativar o cache do objeto de dados, você pode configurar um objeto de dados lógicos ou uma tabela virtual para usar uma tabela de cache gerenciada pelo usuário.

Para usar o tipo de dados de Registro de Data/Hora com Fuso Horário e ativar o armazenamento em cache do objeto de dados para o IBM DB2 ou para o Microsoft SQL Server, defina o formato de data e hora do mapeamento implantado no formato "YYYY-MM-DD HH24:MI:SS". O Serviço de Integração de Dados grava os dados até os segundos.

Tipos de Dados para Tabelas de Cache

O Serviço de Integração de Dados usa dados das tabelas de cache ao processar mapeamentos, consultas de serviços de dados SQL e solicitações de serviços da Web que contêm objetos em cache. Os tipos de dados de tabela de cache esperados pelo Serviço de Integração de Dados podem ser diferentes dos tipos de dados de objetos em cache.

O Gerenciador de Cache do Objeto de Dados cria as tabelas de cache com os tipos de dados esperados pelo Serviço de Integração de Dados. Se você usar tabelas de cache gerenciadas pelo usuário, verifique se as tabelas de cache usam os tipos de dados esperados pelo Serviço de Integração de Dados.

Tipos de Dados de Cache da Tabela Virtual

A seguinte tabela lista os tipos de dados de tabela de cache para tabelas virtuais:

Tipo de Dados da Tabela Virtual	IBM DB2	Microsoft SQL Server	Oracle
Char	Vargraphic Dbclob, para precisão maior que 32672	Nvarchar Ntext, para precisão maior que 4000	Nvarchar2 Nclob, para precisão maior que 2000
Bigint	Bigint	Bigint	Número
Booleano	Número inteiro	Int	Número

Tipo de Dados da Tabela Virtual	IBM DB2	Microsoft SQL Server	Oracle
Data	Registro de data/hora	Datetime2	Registro de data/hora
Double	Double	Flutuante	Registro de data/hora
Decimal	Decimal	Decimal	Número
Int	Número inteiro	Int	Número
Tempo	Registro de data/hora	Datetime2	Registro de data/hora
Registro de data/hora	Registro de data/hora	Datetime2	Registro de data/hora
Varbinary	Blob	Binário Image, para precisão maior que 8000	Brutos Blob, para precisão maior que 2000
Varchar	Vargraphic Dbclob, para precisão maior que 32672	Nvarchar Ntext, para precisão maior que 4000	Nvarchar2 Nclob, para precisão maior que 2000

Tipos de Dados de Cache do Objeto de Dados Lógicos

A seguinte tabela lista os tipos de dados de tabela de cache para os objetos de dados lógicos:

Tipo de Dados do Objeto de Dados Lógicos	DB2	Microsoft SQL Server	Oracle
Bigint	Bigint	Bigint	Número
Binário	Blob	Binário Image, para precisão maior que 8000	Brutos Blob, para precisão maior que 2000
Data e hora	Registro de data/hora	Datetime2	Registro de data/hora
Double	Double	Flutuante	Número
Decimal	Decimal	Decimal	Número
Número inteiro	Número inteiro	Int	Número
String	Vargraphic Dbclob, para precisão maior que 32672	Nvarchar Ntext, para precisão maior que 4000	Nvarchar2 Nclob, para precisão maior que 2000
Texto	Vargraphic Dbclob, para precisão maior que 32672	Nvarchar Ntext, para precisão maior que 4000	Nvarchar2 Nclob, para precisão maior que 2000

Otimização do Cache de Objetos de Dados

O desempenho do cache depende do desempenho do banco de dados de cache e da configuração de objetos em mapeamentos, de serviços de dados SQL e de serviços da Web.

Considere as seguintes soluções para melhorar o desempenho do cache:

Otimizar o banco de dados de cache.

O desempenho ideal para o cache depende da velocidade e do desempenho do banco de dados de cache e do tamanho do cache. Configure o tamanho do cache no banco de dados do cache.

Como o Gerenciador de Cache do Objeto de Dados deve manter o cache antigo para uma operação de atualização, o cache deve ser grande o suficiente para armazenar dois conjuntos de dados. Use a seguinte fórmula para estimar o tamanho do cache mínimo exigido:

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

Por exemplo, você deseja armazenar em cache 20 objetos de dados lógicos e 10 tabelas virtuais. Se o tamanho médio do objeto for 15 MB, o tamanho necessário do cache será $2 * 15 \text{ MB} * (20 + 10) = 900 \text{ MB}$.

Tabelas de cache são somente leitura. Os usuários finais não podem atualizar as tabelas de cache com comandos SQL.

Definir chaves primárias e chaves externas para objetos de dados lógicos.

Quando o Serviço de Integração de Dados gera o cache para objetos de dados lógicos com chaves, ele cria índices. Os índices podem melhorar o desempenho das consultas no banco de dados do cache.

Armazenar em cache objetos de dados lógicos que você associa em um mapeamento.

Quando você associa objetos de dados lógicos em cache, o Serviço de Integração de Dados pode enviar a lógica de transformação de Associador ao banco de dados do cache, mesmo quando os dados de origem são provenientes de bancos de dados diferentes.

Gerar um cache de índice com base nas colunas em um objeto de dados lógicos ou uma tabela virtual.

Configure o Serviço de Integração de Dados para gerar um cache de índices com base nas colunas em objetos de dados lógicos ou tabelas virtuais. O índice pode melhorar o desempenho das consultas no banco de dados do cache.

Otimização do Sistema

Muitas vezes, o desempenho se torna lento porque o mapeamento conta com conexões insuficientes ou com um sistema sobrecarregado de processos do Data Integration Service. Atrasos do sistema também podem ser causados por roteadores, comutadores, protocolos de rede e utilização por diversos usuários.

O acesso lento ao disco em bancos de dados de origem e de destino, em sistemas de arquivos de origem e de destino e em nós no domínio pode diminuir o desempenho do mapeamento. Permita que o administrador do sistema avalie os discos rígidos nas máquinas.

Considere as seguintes soluções para afunilamentos de otimização do sistema:

Aumentar a velocidade da rede.

Conexões de rede lentas podem diminuir o desempenho do mapeamento. Faça com que o administrador do sistema determine se a rede está sendo executada com uma velocidade ideal. Diminua o número de saltos de rede entre o processo do Data Integration Service e bancos de dados.

Usar várias CPUs.

Você pode usar várias CPUs para executar vários mapeamentos em paralelo.

Reduzir a paginação.

Quando um sistema operacional fica sem memória física, ele começa a fazer paginação no disco para liberar memória física. Configure a memória física da máquina de processos do Data Integration Service para minimizar a paginação no disco.

Usar a associação de processador.

Em um ambiente UNIX de vários processadores, o Data Integration Service pode usar uma grande quantidade de recursos do sistema. Use a associação do processador para controlar o uso do processador pelo processo do Serviço de Integração. Além disso, se os bancos de dados de origem e de destino estiverem na mesma máquina, use a associação do processador para limitar os recursos usados pelo banco de dados.

CAPÍTULO 8

Otimização de Serviços de Dados SQL

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Otimização de Serviço de Dados SQL, 70](#)
- [Otimização de Ferramentas de Cliente de Terceiros, 71](#)
- [Níveis do Otimizador do Serviço de Dados SQL, 71](#)
- [Propriedades do Serviço de Dados SQL para Solicitações Simultâneas e de Memória, 75](#)
- [Cache de Conjunto de Resultados para um Serviço de Dados SQL, 76](#)
- [Dados Virtuais Persistentes em Tabelas Temporárias, 78](#)

Visão Geral da Otimização de Serviço de Dados SQL

Você pode otimizar serviço de dados SQL para melhorar o desempenho quando os usuários finais executam consultas SQL nesses serviços usando ferramentas de cliente de terceiros. Se um serviço de dados SQL usar um mapeamento de tabela virtual, você poderá otimizar as origens, as transformações e o mapeamento.

Use as seguintes técnicas de otimização para otimizar um serviço de dados SQL:

- Otimizar ferramentas de cliente de terceiros
- Configurar o nível do otimizador do serviço de dados SQL.
- Configurar propriedades do serviço de dados SQL para simultaneidade e memória para um processo de Integração de Dados.
- Configurar o armazenamento em cache de objetos de dados para o serviço de dados SQL.
- Configurar o armazenamento em cache do conjunto de resultados para o serviço de dados SQL.
- Configure as restrições para tabelas virtuais no serviço de dados SQL.

TÓPICOS RELACIONADOS:

- [“Armazenamento em Cache do Objeto de Dados” na página 65](#)

Otimização de Ferramentas de Cliente de Terceiros

Ferramentas de cliente de terceiros podem afetar o desempenho durante o processamento e a execução de consultas SQL em um serviço de dados SQL. Otimize as ferramentas de cliente de terceiros que os usuários finais podem usar para executar consultas SQL em um serviço de dados SQL.

Considere as seguintes soluções para afunilamentos de ferramentas de cliente de terceiros:

Enviar resultados de consultas grandes para um arquivo em disco.

Uma ferramenta de cliente de terceiros poderá afetar o desempenho se exibir resultados de consultas grandes na janela do console.

Configurar a ferramenta de cliente de terceiros para desabilitar a criptografia.

Uma ferramenta de cliente de terceiros poderá afetar o desempenho se criptografar dados durante a busca ou a exibição de resultados de consultas.

Configurar a ferramenta de cliente de terceiros para buscar previamente um conjunto de linhas.

Uma ferramenta de cliente de terceiros poderá afetar o desempenho se buscar linhas únicas de uma só vez.

Configurar a ferramenta de cliente de terceiros para desabilitar a opção de ler o conteúdo da tabela quando esta é carregada pela primeira vez.

Uma ferramenta de cliente de terceiros poderá afetar o desempenho se as configurações de tipo de dados para os tipos de dados BLOB e CLOB estiverem definidas para ler o conteúdo da tabela quando esta é carregada pela primeira vez, caso os tipos de dados BLOB e CLOB não sejam usados na consulta.

Configurar a ferramenta de cliente de terceiros para usar as configurações padrão de formato e conversão para Data, Hora e Registro de Data/Hora.

Uma ferramenta de cliente de terceiros poderá afetar o desempenho se as configurações de formato e conversão de Data, Hora e Registro de Data/Hora estiverem definidas como um formato especificado pelo usuário, e não como o formato padrão.

Desabilitar a opção de depuração ou defini-la para nenhuma depuração.

Uma ferramenta de cliente de terceiros poderá afetar o desempenho se a opção de depuração para executar a consulta estiver definida como rastreamento. Isso pode diminuir o desempenho, pois a ferramenta de cliente de terceiros grava mais mensagens de log no arquivo de depuração durante o processamento da consulta.

Níveis do Otimizador do Serviço de Dados SQL

O Serviço de Integração de Dados otimiza serviços de dados SQL com base no nível do otimizador que você configura. Configure o nível do otimizador quando quiser que o serviço de dados SQL use um nível de otimizador diferente do normal. Por padrão, cada serviço de dados SQL usa o nível do otimizador normal.

Para compreender como o nível do otimizador cria uma consulta otimizada para um serviço de dados SQL, visualize o plano de consulta para um Serviço de Dados SQL. Quando você visualiza o plano da consulta, a

ferramenta Developer exibe uma representação gráfica da consulta otimizada com base no nível do otimizador e uma representação gráfica da consulta original.

É possível configurar os seguintes níveis do otimizador:

Nenhum

O Serviço de Integração de Dados não aplica nenhuma otimização.

Mínima

O Serviço de Integração de Dados aplica o método de otimização de projeção antecipada.

Normal

O Serviço de Integração de Dados aplica os métodos de otimização de projeção antecipada, seleção antecipada, remoção de ramificação, envio, predicado global e predicado. Normal é o nível de otimização padrão.

Completo

O Serviço de Integração de Dados aplica os métodos de otimização baseada em custos, de projeção antecipada, de seleção antecipada, remoção de ramificação, predicado, envio, semiassociação e dataship-join.

Você pode usar um ou mais dos seguintes métodos para configurar o nível do otimizador para um serviço de dados SQL:

- Configurar o nível do otimizador para visualização de dados de serviço de dados SQL.
- Configure o nível de otimização para serviço de dados SQL implantados.
- Configurar o nível do otimizador na string de conexão de consultas que você executa em um serviço de dados SQL implantado.

Configurando o Nível do Otimizador do Serviço de Dados SQL para Visualização de Dados

Configure o nível do otimizador usado pelo Data Integration Service para executar consultas SQL quando você visualiza a saída de um serviço de dados SQL.

1. Na ferramenta Developer, clique em **Executar > Abrir Caixa de Diálogo de Execução**.
A caixa de diálogo **Executar** é exibida.
2. Clique em **Configuração do Visualizador de Dados**.
3. Clique no botão **Novo**.
4. Insira um nome para a configuração do visualizador de dados.
5. Clique na guia **Avançado**.
6. Selecione um nível do otimizador.
7. Clique em **Aplicar**.
8. Clique em **Fechar**.

A ferramenta Developer cria a configuração do visualizador de dados.

Configurando o Nível do Otimizador para Serviços de Dados SQL Implantados

Configure o nível do otimizador usado pelo Data Integration Services para executar consultas SQL em um serviço de dados SQL implantado. Você pode optar por substituir o nível do otimizador de uma única consulta configurando o nível do otimizador na conexão de serviço de dados SQL.

1. Na ferramenta Administrator, selecione um Data Integration Service.
2. Clique na exibição **Aplicativos**.
3. Expanda o aplicativo que contém o serviço de dados SQL cujo nível do otimizador você deseja configurar.
4. Selecione o serviço de dados SQL e edite a seguinte propriedade:

Propriedade	Descrição
Nível de Otimização	O nível do otimizador que o Serviço de Integração de Dados aplica ao objeto. Insira o valor numérico associado ao nível do otimizador que você deseja configurar. Você pode inserir um dos seguintes valores numéricos: <ul style="list-style-type: none">- 0. O Serviço de Integração de Dados não se aplica à otimização.- 1. O Serviço de Integração de Dados aplica o método de otimização de projeção antecipada.- 2. O Serviço de Integração de Dados aplica os métodos de projeção antecipada, de seleção antecipada e de otimização de envio e predicado.- 3. O Serviço de Integração de Dados aplica os métodos com base em custos, de projeção antecipada, de seleção antecipada e de otimização de envio, predicado e semiassociação.

5. Para substituir nível do otimizador usado pelo Data Integration Services para executar uma consulta, acrescente a seguinte entrada à string de conexão JDBC URL ou ODBC:

```
SQLDataServiceOptions.optimizeLevel= <nível_do_otimizador_numérico>.
```

Plano de Consulta do Serviço de Dados SQL

Ao visualizar o plano da consulta para um serviço de dados SQL, você visualiza a representação gráfica da consulta original e a representação gráfica da consulta otimizada. A representação gráfica descreve como o Data Integration Service processa a consulta. Ela inclui as transformações e a ordem na qual o Data Integration Services processa cada transformação.

A ferramenta Developer usa o nível do otimizador que você define nela para gerar a consulta otimizada. A consulta otimizada exibe a consulta à medida que ela é executada pelo Data Integration Service.

Por exemplo, você deseja consultar a tabela virtual CUSTOMERS em um serviço de dados SQL. Na exibição do **Visualizador de Dados**, você escolhe as definições padrão de configuração do visualizador de dados, o que define como normal o nível do otimizador para a consulta.

Insira a seguinte consulta na exibição do **Visualizador de Dados**:

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

Quando você visualiza o plano de consulta SQL, a ferramenta Developer exibe a seguinte representação gráfica da consulta:



A exibição não otimizada mostra a consulta que você insere. A ferramenta Developer mostra a cláusula WHERE como uma transformação de Filtro e a cláusula ORDER BY como uma transformação de Classificador. A ferramenta Developer usa a transformação de Expressão de passagem para renomear portas.

Quando você visualiza a consulta otimizada, a ferramenta Developer exibe a seguinte representação gráfica da consulta:



A exibição otimizada mostra a consulta que é executada pelo Data Integration Service. Como o nível do otimizador é normal, o Data Integration Service envia a condição de filtro ao objeto de dados de origem. Enviar a condição de filtro melhora o desempenho da consulta, pois reduz o número de linhas lidas pelo Data Integration Service a partir do objeto de dados de origem. De maneira semelhante à consulta não otimizada, a ferramenta Developer exibe a cláusula ORDER BY como uma transformação de Classificador. Ele usa transformações de Expressão de passagem para impor os tipos de dados que você especifica nas transformações lógicas.

Exibindo um Plano de Consulta SQL

Mostre o plano de consulta SQL para exibir uma representação ao estilo de mapeamento da consulta SQL que você insere ao visualizar dados da tabela virtual.

1. Abra um serviço de dados SQL que contenha pelo menos uma tabela virtual.
2. Clique na exibição **Visualizador de Dados**.
3. Insira uma consulta SQL na janela **Entrada**.
4. Como alternativa, selecione uma configuração do visualizador de dados que contenha o nível do otimizador que você deseja aplicar à consulta.
5. Clique em **Mostrar Plano da Consulta**.

A ferramenta Developer exibe o plano de consulta SQL para a consulta inserida na guia **Não Otimizada**.

6. Para exibir a consulta otimizada, clique na guia **Otimizada**.

A ferramenta Developer exibe o plano de consulta SQL otimizado.

Propriedades do Serviço de Dados SQL para Solicitações Simultâneas e de Memória

Para otimizar o desempenho do serviço de dados SQL, configure as propriedades de simultaneidade e de memória do Serviço de Integração de Dados na ferramenta Administrator.

A seguinte tabela descreve a propriedade Memória Máxima por Solicitação para o Módulo de Serviço SQL:

Propriedade	Descrição
Memória Máxima por Solicitação	<p>O comportamento da Memória Máxima por Solicitação depende das seguintes configurações do Serviço de Integração de Dados:</p> <ul style="list-style-type: none">- O serviço executa tarefas em processos locais ou remotos separados ou o tamanho máximo de memória da propriedade do serviço é 0 (padrão). <p>A Memória Máxima por Solicitação é a quantidade máxima de memória, em bytes, que o Serviço de Integração de Dados pode alocar em todas as transformações que usam o modo de cache automático em uma única solicitação. O serviço aloca memória separadamente nas transformações com um tamanho do cache específico. O total de memória usada pela solicitação pode exceder o valor de Máximo de Memória por Solicitação.</p> <ul style="list-style-type: none">- O serviço executa tarefas no processo do Serviço de Integração de Dados e o tamanho máximo de memória da propriedade do serviço é 0 (padrão). <p>A Memória Máxima por Solicitação é a quantidade máxima de memória, em bytes, que o Serviço de Integração de Dados pode alocar em uma única solicitação. O total de memória usada pela solicitação não pode exceder o valor de Máximo de Memória por Solicitação.</p> <p>O padrão é 50.000.000.</p>

A seguinte tabela descreve a propriedade Tamanho Máximo do Heap para o processo do Serviço de Integração de Dados:

Propriedade	Descrição
Tamanho Máximo do Heap	<p>Quantidade de RAM alocada na Máquina Virtual Java (JVM) que executa o Serviço de Integração de Dados. Use essa propriedade para melhorar o desempenho. Acrescente uma das seguintes letras ao valor para especificar as unidades:</p> <ul style="list-style-type: none">- b para bytes.- k para kilobytes.- m para megabytes.- g para gigabytes. <p>O padrão é 640 megabytes.</p> <p>Nota: Considere aumentar o tamanho do heap quando o Serviço de Integração de Dados necessitar processar grandes quantidades de dados.</p>

A seguinte tabela descreve as propriedades SQL para o processo do Serviço de Integração de Dados:

Propriedade	Descrição
Nº Máximo de Conexões Simultâneas	Limita o número de conexões de banco de dados que o Serviço de Integração de Dados pode fazer para os serviços de dados SQL. O padrão é 100.

A seguinte tabela descreve as opções de execução do Serviço de Integração de Dados:

Propriedade	Descrição
Tamanho Máximo do Pool de Execução	<p>Número máximo de trabalhos que cada processo do Serviço de Integração de Dados pode executar simultaneamente. Os trabalhos incluem visualizações de dados, mapeamentos, trabalhos de criação de perfil, consultas SQL e solicitações de serviço da Web. Por exemplo, uma grade do Serviço de Integração de Dados inclui três processos de serviço em execução. Se você definir o valor como 10, cada processo do Serviço de Integração de Dados poderá executar até 10 trabalhos simultaneamente. Um total de 30 trabalhos pode ser executado simultaneamente na grade. O padrão é 10.</p> <p>Nota: Quando você aumenta o valor do tamanho do pool, o Serviço de Integração de Dados usa mais recursos de hardware, como CPU, memória e E/S do sistema. Defina esse valor com base nos recursos disponíveis na máquina. Por exemplo, considere o número de CPUs na máquina que hospeda o Serviço de Integração de Dados e a quantidade de memória que está disponível para o Serviço de Integração de Dados.</p>
Tamanho Máximo de Memória	<p>Quantidade máxima de memória, em bytes, que o Serviço de Integração de Dados pode alocar para a execução de todas as solicitações simultaneamente quando o serviço executa tarefas em um processo do Serviço de Integração de Dados. Quando o Serviço de Integração de Dados executa tarefas em processos locais ou remotos separados, o serviço ignora esse valor. Se você não quiser limitar a quantidade de memória que o Serviço de Integração de Dados pode alocar, defina essa propriedade como 0.</p> <p>Se o valor for maior que 0, o Serviço de Integração de Dados usará a propriedade para calcular o máximo de memória total permitido para a execução de todas as solicitações simultaneamente. O Serviço de Integração de Dados calcula o máximo de memória total da seguinte forma:</p> <p>Tamanho Máximo de Memória + Tamanho Máximo do Heap + memória necessária para carregar componentes do programa</p> <p>O padrão é 0.</p> <p>Nota: Se você executar perfis ou mapeamentos de qualidade de dados, defina essa propriedade como 0.</p>

Cache de Conjunto de Resultados para um Serviço de Dados SQL

Quando você configura o cache do conjunto de resultados, o Serviço de Integração de Dados armazena em cache os resultados do processo do DTM associado a cada consulta de serviço de dados SQL e solicitação de serviço da Web. O Serviço de Integração de Dados armazena os resultados para o período de expiração que você configurar. Quando um cliente faz a mesma consulta antes da expiração do cache, o Serviço de Integração de Dados retorna os resultados em cache.

Considere a seguinte solução para afunilamentos de cache de conjunto de resultados:

Configurar o cache do conjunto de resultados para um serviço de dados SQL.

O cache do conjunto de resultados permite que o Serviço de Integração de Dados use resultados em cache para consultas de serviço de dados SQL. Os usuários que executam consultas idênticas em um curto período de tempo podem querer usar o cache de conjunto de resultados para diminuir o tempo de execução de consultas idênticas.

Quando você habilita o Serviço de Integração de Dados para usar resultados em cache, o desempenho do serviço de dados melhora. No entanto, para melhorar ainda o tempo de processamento do serviço de dados para consultas idênticas, aloque espaço suficiente para armazenar o cache na memória. Ao configurar a quantidade de memória de cache de forma que ela seja igual ou superior à quantidade

necessária para armazenar os resultados em cache, você melhora o desempenho reduzindo a sobrecarga de E/S do sistema. Quando o Serviço de Integração de Dados grava arquivos de cache no disco, o tempo de processamento dos serviços de dados aumenta devido à sobrecarga de E/S do sistema.

Propriedades do Cache de Conjunto de Resultados do Serviço de Dados SQL

Para melhorar o desempenho, você pode configurar as propriedades do cache de conjunto de resultados para um Data Integration Service. Também pode configurar por quantos milissegundos o cache do conjunto de resultados fica disponível para uso por um serviço de dados SQL.

A tabela a seguir descreve as propriedades do cache de conjunto de resultados para o Data Integration Service:

Propriedade	Descrição
Prefixo do Nome do Arquivo	O prefixo dos nomes de todos os arquivos de cache do conjunto de resultados armazenados em disco. O padrão é RSCACHE.
Ativar Criptografia	Indica se os arquivos de cache do conjunto de resultados estão criptografados usando o padrão AES de 128 bits. Os valores válidos são verdadeiro ou falso. O padrão é True.

A tabela a seguir descreve a propriedade que configura o número de milissegundos durante os quais o cache de conjunto de resultados está disponível para o serviço de dados SQL:

Propriedade	Descrição
Período de Expiração do Cache de Conjunto de Resultados	O número de milissegundos durante os quais o cache de conjunto de resultados está disponível para uso. Se for definido como -1, o cache não expirará nunca. Se for definido como 0, o cache de conjunto de resultados ficará desabilitado. Alterações no período de expiração não se aplicam a caches existentes. Se você quiser que todos os caches usem o mesmo período de expiração, limpe o cache de conjunto de resultados depois de alterar o período de expiração. O padrão é 0.

Habilitando o Armazenamento em Cache do Conjunto de Resultados para um Serviço de Dados SQL

Para usar os resultados em cache para consultas de serviço de dados SQL idênticas, configure o Data Integration Service para usar o armazenamento em cache de conjunto de resultados.

1. Na ferramenta Administrator, selecione um Data Integration Service.
2. Clique na exibição **Processo** para configurar as propriedades do cache do conjunto de resultados.
3. Clique na exibição **Aplicativo** e clique no serviço de dados SQL para configurar a propriedade Expiração do Cache de Conjunto de Resultados.

Dados Virtuais Persistentes em Tabelas Temporárias

Uma tabela temporária é uma tabela em um banco de dados relacional que armazena dados temporários e intermediários. Geralmente as consultas complexas necessitam de armazenamento para grandes quantidades de dados intermediários, como informações de associações. Quando você implementa tabelas temporárias, as ferramentas de inteligência comercial podem recuperar esses dados da tabela temporária, em vez do serviço de dados SQL. Isso resulta em um aumento no desempenho.

As tabelas temporárias também oferecem uma maior segurança de duas formas. Primeiro, somente o usuário da sessão ativa pode acessar as tabelas. Além disso, as tabelas persistem enquanto a sessão está ativa e o banco de dados descarta tabelas quando a conexão é encerrada.

Implementação da Tabela Temporária

Você pode usar tabelas temporárias para melhorar o desempenho das consultas grandes e complexas. As tabelas temporárias melhoram o desempenho porque as consultas de tabelas temporárias em um banco de dados relacional são mais rápidas que as consultas repetidas do serviço de dados SQL para o mesmo conjunto de dados.

A implementação de tabelas temporárias para o aprimoramento do desempenho requer ações do administrador da Informática e um desenvolvedor de ferramentas de inteligência comercial.

Primeiro, o administrador da Informática cria uma conexão de banco de dados relacional e configura o Serviço de Integração de Dados para usar a conexão.

Em seguida, o desenvolvedor de uma ferramenta de inteligência comercial (por exemplo, IBM Cognos ou SAP Business Objects) cria uma conexão entre a ferramenta de inteligência comercial e o serviço de dados SQL da Informática. A conexão usa o driver ODBC ou JDBC da Informática.

Quando essas conexões estão ativas, a ferramenta de inteligência comercial pode criar e usar tabelas temporárias para processar grandes quantidades de dados intermediários.

CAPÍTULO 9

Otimização de Serviços da Web

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Otimização de Serviços da Web, 79](#)
- [Otimizar Solicitações HTTP, 80](#)
- [Compactação de Mensagens de Serviços da Web, 80](#)
- [Nível do Otimizador de Serviços da Web, 80](#)
- [Propriedades dos Serviços da Web para Solicitações Simultâneas e de Memória , 82](#)
- [Propriedade de Serviço da Web para Configurar uma Instância do DTM Ativo, 85](#)
- [Armazenamento em Cache de Conjunto de Resultados do Serviço da Web, 85](#)
- [Gerenciamento de Logs de Serviços da Web, 86](#)

Visão Geral da Otimização de Serviços da Web

Você pode otimizar serviços da Web para melhorar o desempenho quando o Data Integration Service executa solicitações de serviços da Web. Ajuste o Data Integration Service para gerenciar memória e lidar com solicitações de serviços da Web simultâneas. Para melhorar o desempenho de serviços da Web, use a compactação de mensagens de serviços da Web, otimize solicitações HTTP, configure o cache de conjunto de resultados e objetos de dados e configure níveis de log de erros.

Use as seguintes técnicas de otimização para otimizar um serviço da Web:

- Otimizar solicitações HTTP.
- Compactar mensagens de serviços da Web.
- Configurar o nível do otimizador de serviços da Web.
- Configurar propriedades de serviços da Web para simultaneidade e memória para um processo de Integração de Dados.
- Configurar o Data Integration Service para manter um processo do DTM ativo de forma que ele possa processar mais de uma solicitação de serviço da Web.
- Configurar o armazenamento em cache de objetos de dados para o serviço da Web.
- Configurar o armazenamento em cache do conjunto de resultados para os serviços da Web.
- Configurar os níveis de log de erros de tempo de execução dos serviços da Web.

TÓPICOS RELACIONADOS:

- [“Armazenamento em Cache do Objeto de Dados” na página 65](#)

Otimizar Solicitações HTTP

Otimize solicitações HTTP para reduzir o número de solicitações para o servidor da Web.

Considere as seguintes soluções para afunilamentos de solicitações HTTP:

Reduzir o tempo limite de soquetes HTTP para o cliente de serviços da Web.

O tempo limite de soquetes define por quanto tempo o cliente aguarda antes de atingir o tempo limite da solicitação HTTP. O cliente de serviços da Web poderá travar se o valor de tempo limite de soquetes for grande.

Compactação de Mensagens de Serviços da Web

Você pode otimizar o desempenho de serviços da Web compactando mensagens da Web grandes que são transmitidas para e a partir de provedores.

Considere a seguinte solução para afunilamentos de mensagens de serviços da Web:

Habilitar a compactação de mensagens SOAP para um cliente de serviços da Web.

A compactação de mensagens SOAP permite que o serviço da Web seja compactado para receber mensagens compactadas do cliente de serviços da Web. O serviço da Web pode aceitar uma mensagem SOAP com compactação GZip a partir de um cliente de serviços da Web.

Quando o Data Integration Service receber a resposta do serviço da Web, ele verificará o cabeçalho HTTP de Codificação de Conteúdo na mensagem SOAP e decodificará essa mensagem.

Nível do Otimizador de Serviços da Web

O Serviço de Integração de Dados otimiza serviços da Web com base no nível do otimizador que você configura. Configure o nível do otimizador quando quiser que o serviço da Web use um nível de otimizador diferente do normal. Por padrão, cada serviço da Web usa o nível do otimizador normal.

Você pode escolher um dos seguintes níveis de otimizador:

Nenhum

O Serviço de Integração de Dados não aplica nenhuma otimização.

Mínima

O Serviço de Integração de Dados aplica o método de otimização de projeção antecipada.

Normal

O Serviço de Integração de Dados aplica os métodos de otimização de projeção antecipada, seleção antecipada, remoção de ramificação, envio, predicado global e predicado. Normal é o nível de otimização padrão.

Completo

O Serviço de Integração de Dados aplica os métodos de otimização baseada em custos, de projeção antecipada, de seleção antecipada, remoção de ramificação, predicado, envio, semiassociação e dataship-join.

Você pode usar um ou mais dos seguintes métodos para configurar o nível do otimizador para um serviço da Web:

- Configurar o nível do otimizador para visualização de dados de serviços da Web antes de implantá-los em um Serviço de Integração de Dados.
- Configurar o nível de otimização para serviços da Web implantados que são executados em um Serviço de Integração de Dados específico.
- Configurar o nível do otimizador no cabeçalho da solicitação de serviço da Web para um serviço da Web implantado.

Configurando o Nível do Otimizador do Serviço da Web para Visualização de Dados

Configure o nível do otimizador usado pelo Data Integration Services para visualizar a saída de um serviço da Web.

1. Na ferramenta Developer, clique em **Executar** > **Abrir Caixa de Diálogo de Execução**.
A caixa de diálogo **Executar** é exibida.
2. Clique em **Configuração do Serviço da Web**.
3. Clique no botão **Novo**.
4. Insira um nome para a configuração de serviço da Web.
5. Clique na guia **Avançado**.
6. Selecione um nível do otimizador.
7. Clique em **Aplicar**.
8. Clique em **Fechar**.

A ferramenta Developer cria a configuração do serviço da Web.

Ao executar o visualizador de dados para visualizar a saída de uma operação de mapeamento, selecione a configuração de serviço da Web que inclui o nível de otimizador que você deseja usar.

Configurando o Nível do Otimizador para Serviços da Web Implantados

Configure o nível do otimizador usado pelo Data Integration Services para executar um serviço da Web implantado. Você pode optar por substituir o nível do otimizador para uma única solicitação configurando o nível do otimizador no cabeçalho HTTP da solicitação SOAP de serviços da Web.

1. Na ferramenta Administrator, selecione um Data Integration Service.
2. Clique na exibição **Aplicativos**.
3. Expanda o aplicativo que contém o serviço da Web cujo nível do otimizador você deseja configurar.

4. Selecione o serviço da Web e edite a seguinte propriedade:

Propriedade	Descrição
Nível de Otimização	<p>O nível do otimizador que o Serviço de Integração de Dados aplica ao objeto. Insira o valor numérico associado ao nível do otimizador que você deseja configurar. Você pode inserir um dos seguintes valores numéricos:</p> <ul style="list-style-type: none">- 0. O Serviço de Integração de Dados não se aplica à otimização.- 1. O Serviço de Integração de Dados aplica o método de otimização de projeção antecipada.- 2. O Serviço de Integração de Dados aplica os métodos de projeção antecipada, de seleção antecipada e de otimização de envio e predicado.- 3. O Serviço de Integração de Dados aplica os métodos com base em custos, de projeção antecipada, de seleção antecipada e de otimização de envio, predicado e semiassociação.

5. Para substituir o nível de otimização de serviços da Web para uma solicitação de serviço da Web, inclua a seguinte entrada no cabeçalho HTTP da solicitação SOAP desse serviço da Web:

```
WebServiceOptions.optimizeLevel= <nível_do_otimizador_numérico>.
```

Propriedades dos Serviços da Web para Solicitações Simultâneas e de Memória

Para otimizar o desempenho de serviços da Web, configure as propriedades de simultaneidade e de memória para o Serviço de Integração de Dados e para cada serviço da Web na ferramenta Administrator.

A seguinte tabela descreve a propriedade Memória Máxima por Solicitação para o Módulo do Serviço da Web:

Propriedade	Descrição
Memória Máxima por Solicitação	<p>O comportamento da Memória Máxima por Solicitação depende das seguintes configurações do Serviço de Integração de Dados:</p> <ul style="list-style-type: none">- O serviço executa tarefas em processos locais ou remotos separados ou o tamanho máximo de memória da propriedade do serviço é 0 (padrão). A Memória Máxima por Solicitação é a quantidade máxima de memória, em bytes, que o Serviço de Integração de Dados pode alocar em todas as transformações que usam o modo de cache automático em uma única solicitação. O serviço aloca memória separadamente nas transformações com um tamanho do cache específico. O total de memória usada pela solicitação pode exceder o valor de Máximo de Memória por Solicitação.- O serviço executa tarefas no processo do Serviço de Integração de Dados e o tamanho máximo de memória da propriedade do serviço é 0 (padrão). A Memória Máxima por Solicitação é a quantidade máxima de memória, em bytes, que o Serviço de Integração de Dados pode alocar em uma única solicitação. O total de memória usada pela solicitação não pode exceder o valor de Máximo de Memória por Solicitação. O padrão é 50.000.000.

A seguinte tabela descreve as opções de execução do Serviço de Integração de Dados:

Propriedade	Descrição
Tamanho Máximo do Pool de Execução	<p>Número máximo de trabalhos que cada processo do Serviço de Integração de Dados pode executar simultaneamente. Os trabalhos incluem visualizações de dados, mapeamentos, trabalhos de criação de perfil, consultas SQL e solicitações de serviço da Web. Por exemplo, uma grade do Serviço de Integração de Dados inclui três processos de serviço em execução. Se você definir o valor como 10, cada processo do Serviço de Integração de Dados poderá executar até 10 trabalhos simultaneamente. Um total de 30 trabalhos pode ser executado simultaneamente na grade. O padrão é 10.</p> <p>Nota: Quando você aumenta o valor do tamanho do pool, o Serviço de Integração de Dados usa mais recursos de hardware, como CPU, memória e E/S do sistema. Defina esse valor com base nos recursos disponíveis na máquina. Por exemplo, considere o número de CPUs na máquina que hospeda o Serviço de Integração de Dados e a quantidade de memória que está disponível para o Serviço de Integração de Dados.</p>
Tamanho Máximo de Memória	<p>Quantidade máxima de memória, em bytes, que o Serviço de Integração de Dados pode alocar para a execução de todas as solicitações simultaneamente quando o serviço executa tarefas em um processo do Serviço de Integração de Dados. Quando o Serviço de Integração de Dados executa tarefas em processos locais ou remotos separados, o serviço ignora esse valor. Se você não quiser limitar a quantidade de memória que o Serviço de Integração de Dados pode alocar, defina essa propriedade como 0.</p> <p>Se o valor for maior que 0, o Serviço de Integração de Dados usará a propriedade para calcular o máximo de memória total permitido para a execução de todas as solicitações simultaneamente. O Serviço de Integração de Dados calcula o máximo de memória total da seguinte forma:</p> <p>Tamanho Máximo de Memória + Tamanho Máximo do Heap + memória necessária para carregar componentes do programa</p> <p>O padrão é 0.</p> <p>Nota: Se você executar perfis ou mapeamentos de qualidade de dados, defina essa propriedade como 0.</p>

A seguinte tabela descreve as propriedades de configuração HTTP para o processo do Serviço de Integração de Dados:

Propriedade	Descrição
Número Máximo de Solicitações de Registros Acumulados	<p>Número máximo de conexões HTTP ou HTTPS que podem aguardar em fila para esse processo do Serviço de Integração de Dados. O padrão é 100.</p>
Número Máximo de Solicitações Simultâneas	<p>Número máximo de conexões HTTP ou HTTPS para esse processo do Serviço de Integração de Dados. O mínimo é 4. O padrão é 200.</p> <p>Nota: Para um serviço da Web, essa propriedade afeta o número de solicitações de serviços da Web que o Serviço de Integração de Dados aceita antes de enviar as solicitações ao registro acumulado do Serviço de Integração de Dados.</p>

A seguinte tabela descreve a propriedade Tamanho Máximo do Heap que você pode configurar para o processo do Serviço de Integração de Dados:

Propriedade	Descrição
Tamanho Máximo do Heap	Quantidade de RAM alocada na Máquina Virtual Java (JVM) que executa o Serviço de Integração de Dados. Use essa propriedade para melhorar o desempenho. Acrescente uma das seguintes letras ao valor para especificar as unidades: <ul style="list-style-type: none">- b para bytes.- k para kilobytes.- m para megabytes.- g para gigabytes. O padrão é 640 megabytes. Nota: Considere aumentar o tamanho do heap quando o Serviço de Integração de Dados necessitar processar grandes quantidades de dados.

Exemplo de Configuração do Serviço de Integração de Dados para Solicitações de Serviços da Web Simultâneas

Ao configurar como o Serviço de Integração de Dados processa solicitações de serviços da Web simultâneas, verifique se o valor para o número máximo de solicitações simultâneas é o mesmo para o serviço da Web e o processo do Serviço de Integração de Dados.

Por exemplo, na configuração a seguir, o Serviço de Integração de Dados aceita 200 solicitações HTTP simultâneas, mas apenas 10 solicitações simultâneas de serviços da Web:

Tipo de Propriedade	Nome da Propriedade	Configuração
Processo do Serviço de Integração de Dados	Número Máximo de Solicitações Simultâneas	200
Processo do Serviço de Integração de Dados	Número Máximo de Solicitações de Registros Acumulados	500
Serviço de Integração de Dados	Tamanho Máximo do Pool de Execução	100
Serviço da Web	Número Máximo de Solicitações Simultâneas	10

Quando o Serviço de Integração de Dados recebe 20 solicitações de serviços da Web, 10 delas apresentam falhas, pois o serviço da Web só pode aceitar 10 solicitações simultâneas.

Para evitar a falha de solicitações de serviço da Web quando o serviço da Web atingir seu número máximo de solicitações simultâneas, configure o mesmo valor máximo para o processo do Serviço de Integração de Dados e o serviço da Web. Quando o número de solicitações enviadas ao Serviço de Integração de Dados excede o valor máximo de solicitações simultâneas, as solicitações adicionais permanecem no registro acumulado até o processo do Serviço de Integração de Dados ficar disponível para processar as solicitações.

Propriedade de Serviço da Web para Configurar uma Instância do DTM Ativo

Para melhorar o desempenho, você pode configurar o Serviço de Integração de Dados para manter uma instância do DTM ativa de forma que ela possa processar mais de uma solicitação de serviço da Web. Você pode configurar a propriedade Tempo de Atividade do DTM para o Serviço de Integração de Dados na ferramenta Administrator.

A tabela a seguir descreve a propriedade Tempo de Atividade do DTM:

Propriedade	Descrição
Tempo de Atividade do DTM	<p>Número de milissegundos pelo qual a instância do DTM permanece aberta após concluir a última solicitação. Solicitações de serviço da Web que são emitidas com a mesma operação podem reutilizar a instância aberta. Use o tempo de atividade para melhorar o desempenho quando o tempo necessário para processar a solicitação for pequeno em comparação com o tempo de inicialização da instância do DTM. Se a solicitação falhar, a instância do DTM será encerrada.</p> <p>O padrão é 5000.</p> <p>Nota: A capacidade de usar um processo do DTM existente aumenta o desempenho. O DIS exige recursos adicionais para iniciar um processo do DTM para cada solicitação. Manter o DTM ativo consome memória. Portanto, os usuários devem considerar o consumo da memória ao configurarem essa opção.</p>

Armazenamento em Cache de Conjunto de Resultados do Serviço da Web

Quando você configura o armazenamento em cache de conjunto de resultados, o Data Integration Service armazena em cache os resultados do processo do DTM associado a cada solicitação de serviço da Web. O Data Integration Service armazena os resultados para o período de expiração que você configurar. Quando um cliente externo faz a mesma solicitação antes da expiração do cache, o Data Integration Service retorna os resultados em cache.

Considere a seguinte solução para afunilamentos de cache de conjunto de resultados:

Configurar o cache do conjunto de resultados para um serviço da Web.

O armazenamento em cache de conjunto de resultados permite que o Data Integration Service use os resultados em cache para solicitações de serviços da Web. Os usuários que executam consultas idênticas em um curto período podem querer usar o armazenamento em cache de conjunto de resultados para reduzir o tempo de execução de consultas idênticas.

O Data Integration Service armazena o cache do conjunto de resultados para serviços da Web por usuário quando o serviço da Web usa WSSecurity. O Data Integration Service armazena o cache com base no nome do usuário que é fornecido no token de nome de usuário da solicitação de serviço da Web. Quando o Data Integration Service armazena os resultados em cache por usuário, o Data Integration Service retorna os resultados em cache somente para o usuário que enviou a solicitação de serviço da Web.

Habilitando o Armazenamento em Cache de Conjunto de Resultados para um Serviço da Web

Para usar os resultados em cache para solicitações de serviços da Web idênticas, configure o Data Integration Service para usar o armazenamento em cache de conjunto de resultados.

1. Na ferramenta Administrator, selecione um Data Integration Service.
2. Clique na exibição **Processo** para configurar as propriedades do cache do conjunto de resultados.
3. Clique na exibição **Aplicativo**, clique no serviço da Web e depois clique na operação para configurar o período de expiração do cache nas propriedades da operação do serviço da Web. Se quiser que o Data Integration Service armazene os resultados em cache por usuário, habilite a especificação WS-Security nas propriedades do serviço da Web.
4. Para desabilitar o armazenamento em cache do conjunto de resultados para uma solicitação de serviço da Web quando a operação de serviço da Web está configurada para armazenar em cache o conjunto de resultados, inclua a seguinte sintaxe no cabeçalho HTTP da solicitação SOAP:

```
WebServiceOptions.disableResultSetCache=true
```

Gerenciamento de Logs de Serviços da Web

O desempenho de E/S do sistema pode diminuir quando o Data Integration Service grava e mantém muitos arquivos de log. O Data Integration Service gera logs de tempo de execução de serviços da Web com base no nível de rastreamento que você configura. Considere o gerenciamento do número de arquivos de log que são gravados e mantidos pelo Data Integration Service.

Considere as seguintes soluções para afunilamentos de logs de serviços da Web:

Definir como inativo o nível de rastreamento do serviço da Web.

Ao configurar propriedades de serviços da Web para um serviço da Web implantado, você pode especificar o nível de rastreamento do log. O nível de rastreamento determina os tipos de logs que o Data Integration Service grava na localização dos logs de tempo de execução. O nível de rastreamento padrão do serviço da Web é INFO. Quando o nível de rastreamento está definido como FINEST ou ALL, o desempenho pode diminuir porque o Data Integration Service grava mais logs no arquivo de log. O impacto no desempenho quando o nível de rastreamento é definido como FINEST ou ALL é maior quando o serviço da Web usa HTTPS e a especificação WS-Security.

Arquivar arquivos de log que não são mais necessários.

A E/S do sistema é afetada com o armazenamento de muitos arquivos de log. Por padrão, o Data Integration Service grava os logs de tempo de execução de serviços da Web no seguinte diretório:

```
<InformaticaInstallationDir>/tomcat/bin/disLogs/ws
```

Nota: Se você excluir a pasta ws para esvaziar os logs, deverá recriar essa pasta. Pare o Data Integration Service antes de excluir e recriar a pasta ws.

CAPÍTULO 10

Otimização de Conexões

Este capítulo inclui os seguintes tópicos:

- [Visão Geral da Otimização de Conexões, 87](#)
- [Pool de Conexões, 87](#)
- [Tamanho dos Pacotes de Rede do Banco de Dados, 88](#)

Visão Geral da Otimização de Conexões

Você pode otimizar conexões para melhorar o desempenho. É possível gerenciar o pool de instâncias de conexão inativas para uma conexão de banco de dados. Além disso, você pode aumentar o tamanho dos pacotes de rede para permitir que pacotes de dados maiores atravessem a rede de uma vez.

Use as seguintes técnicas para otimizar conexões:

- Otimizar o pool de conexões.
- Otimizar o tamanho dos pacotes de rede do banco de dados.

Pool de Conexões

O pool de conexões é uma estrutura para armazenar em cache informações de conexão de banco de dados que são usadas pelo Serviço de Integração de Dados. Essa estrutura aumenta o desempenho através da reutilização de informações de conexão em cache.

Considere a seguinte solução para afunilamentos de conexões:

Habilitar o pool de conexões para uma conexão de banco de dados.

Habilitar o pool de conexões para otimizar o desempenho da conexão. É possível gerenciar as instâncias de conexão inativas para uma conexão de banco de dados. O pool de conexões retém instâncias de conexão inativas com base nas propriedades do pool que você configura. É possível ajustar o número máximo e o número mínimo de conexões inativas, bem como o tempo de espera máximo por uma conexão inativa.

Propriedades do Pool em Objetos de Conexão

Você pode editar as propriedades do pool de conexões na exibição **Pool** de uma conexão de banco de dados.

O número de bibliotecas do pool conexão depende do número de processos do Serviço de Integração de Dados ou do DTM em execução. Cada processo do Serviço de Integração de Dados ou do DTM mantém sua própria biblioteca de pool de conexões. Os valores das propriedades do pool destinam-se a cada biblioteca de pool de conexões.

Por exemplo, se você definir o máximo de conexões como 15, cada biblioteca de pools de conexão poderá ter no máximo 15 conexões inativas no pool. Se o Serviço de Integração de Dados executar os trabalhos em processos locais separados e três processos do DTM estiverem em execução, você poderá ter um máximo de 45 instâncias conexão ociosas.

Para reduzir o número total de instâncias de conexões inativas, defina o número mínimo de conexões como 0 e reduza o tempo máximo de inatividade para cada conexão de banco de dados.

A lista a seguir descreve as propriedades do pool de conexões de banco de dados que podem ser editadas na exibição **Pool** para uma conexão de banco de dados:

Ativar Pool de Conexões

Ativa o pool de conexões. Quando você ativa o pool de conexões, cada pool de conexão retém as instâncias de conexões inativas na memória. Para excluir o pool de conexões inativas, você deve reiniciar o Serviço de Integração de Dados.

Se o pool de conexões for desativado, o processo do DTM ou o processo do Serviço de Integração de Dados interromperá todas as atividades de pool. O processo do DTM ou o processo do Serviço de Integração de Dados cria uma instância de conexão sempre que processa um trabalho. Ele descarta a instância ao concluir o processamento do trabalho.

O padrão é ativado para conexões DB2 para i5/OS, DB2 para z/OS, IBM DB2, Microsoft SQL Server, Oracle e ODBC. O padrão é desativado para conexões Adabas, IMS, Sequenciais e VSAM.

Nº Mínimo de Conexões

O número mínimo de instâncias de conexões inativas que um pool mantém para uma conexão de banco de dados depois que o tempo de inatividade máximo é atingido. Defina esse valor como igual ou menor que o número máximo de instâncias de conexões inativas. O padrão é 0.

Nº Máximo de Conexões

O número máximo de instâncias de conexões inativas que um pool mantém para uma conexão de banco de dados antes que o tempo máximo de inatividade seja atingido. Defina esse valor como sendo maior que o número mínimo de instâncias de conexões inativas. O padrão é 15.

Tempo de Inatividade Máximo

O número de segundos pelo qual uma instância de conexão que excede o número mínimo de instâncias de conexão pode permanecer inativa antes de ser descartada pelo pool de conexões. O pool de conexões ignora o tempo ocioso quando a instância de conexão não excede o número mínimo de instâncias de conexões inativas. O padrão é 120.

Tamanho dos Pacotes de Rede do Banco de Dados

Se você ler ou gravar em destinos Oracle, Sybase ASE ou Microsoft SQL Server, poderá melhorar o desempenho aumentando o tamanho dos pacotes de rede com base no banco de dados usado para leitura

ou gravação. Aumentar o tamanho dos pacotes de rede permite que pacotes de dados maiores atravessem a rede de uma vez.

Considere as seguintes soluções para afunilamentos de tamanho dos pacotes de rede do banco de dados:

Aumentar o tamanho dos pacotes de rede de banco de dados para um banco de dados Oracle.

Você pode aumentar o tamanho dos pacotes de rede do servidor de banco de dados em `listener.ora` e em `tnsnames.ora`. Consulte a documentação do seu banco de dados para obter mais informações sobre como aumentar o tamanho dos pacotes, se necessário.

Aumentar o tamanho dos pacotes de rede de banco de dados para um banco de dados Sybase ASE.

Consulte a documentação do seu banco de dados para obter informações sobre como aumentar o tamanho dos pacotes. Você também deve alterar o tamanho dos pacotes para o Sybase ASE no objeto de conexão relacional do Data Integration Service para refletir o tamanho dos pacotes do servidor de banco de dados.

Aumentar o tamanho dos pacotes de rede de banco de dados para um banco de dados Microsoft SQL Server.

Consulte a documentação do seu banco de dados para obter informações sobre como aumentar o tamanho dos pacotes. Você também deve alterar o tamanho dos pacotes para o Microsoft SQL Server no objeto de conexão relacional do Data Integration Service para refletir o tamanho dos pacotes do servidor de banco de dados.

ÍNDICE

A

- afunilamentos
 - no UNIX [12](#)
 - no Windows [11](#)
- ajuste de desempenho
 - método de otimização dataship-join [46](#)
 - método de otimização de ajuste de ramificação [48](#)
 - método de otimização de empilhamento [49](#)
 - método de otimização de seleção antecipada [48](#)
- ajuste do desempenho
 - método de otimização baseada em custos [45](#)
 - método de otimização de envio [49](#)
 - método de otimização de predicado [44](#)
 - método de otimização de predicado global [48](#)
 - método de otimização de projeção antecipada [43](#)
 - método de otimização de semi-associação [46](#)
 - métodos de otimização [42](#)
 - níveis de otimização [42](#)
- alocação de memória
 - Instância ativa do DTM [85](#)
 - Serviço da Web [85](#)
 - Serviço de Dados SQL [75](#)
 - solicitações simultâneas [75](#)
- arquivos simples
 - otimizando destinos para particionamento [56](#)
 - otimizando origens para particionamento [56](#)

B

- bancos de dados
 - otimizando destinos para particionamento [57](#)
 - otimizando origens para particionamento [57](#)

C

- cache de conjunto de resultados
 - habilitando o cache do conjunto de resultados para um serviço de dados SQL [77](#)
 - propriedades do cache de conjunto de resultados [77](#)
- cache de objeto de dados
 - cache de índice [66](#)
 - configurando [66](#)
 - descrição [66](#)
 - otimização [68](#)
 - tabelas gerenciadas pelo usuário [66](#)
 - tipos de dados de tabela [66](#)
- cache de transformações
 - otimização de transformações [35](#)
- cache do conjunto de resultados de serviço de dados SQL
 - Serviço de Integração de Dados [76](#)
- cache do conjunto de resultados de serviços da Web
 - Data Integration Service [85](#)

- cargas em massa
 - otimização de destinos [15](#)
- compactação de mensagens de Serviços da Web
 - otimização de Serviços da Web [80](#)

D

- Data Integration Service
 - cache do conjunto de resultados de serviços da Web [85](#)
- destino de arquivo simples
 - otimização de destinos [14](#)
- dicas
 - exibição Consulta [21](#)
- dicas de banco de dados
 - inserindo na ferramenta Developer [21](#)
- dicas de SQL
 - inserindo na ferramenta Developer [21](#)
- drivers JDBC
 - otimização de tempo de execução [71](#)

E

- efeitos colaterais
 - descrição [36](#)
 - Transformação de Consumidor de Serviço da Web [38](#)
 - transformação SQL [34](#)
- eliminação de erros de transformação
 - otimização de transformações [36](#)
- estatísticas de monitoramento
 - otimização de tempo de execução [63](#)
- exibição Consulta
 - configurando dicas [21](#)

F

- ferramentas de cliente de terceiros
 - otimização de tempo de execução [71](#)
- filtros condicionais
 - otimização de origens [19](#)

G

- gerenciamento de logs de Serviços da Web
 - nível de rastreamento de erros [86](#)

H

- habilitando o cache do conjunto de resultados para um serviço de dados SQL
 - cache de conjunto de resultados [77](#)

I

Instância ativa do DTM
Serviço da Web [85](#)
intervalos de pontos de verificação do banco de dados
otimização de destinos [15](#)

L

leitura de passagem única
otimização de mapeamentos [51](#)

M

mapeamentos
método de otimização de predicado [44](#)
método de otimização de predicado global [48](#)
métodos de otimização [42](#)
otimização particionada [54](#)
método de otimização de empilhamento
empilhamento de origem [51](#)
empilhamento total [50](#)

N

níveis de otimização
descrição [42](#)
nível de otimização mínimo
descrição [42](#)
nível de otimização normal
descrição [42](#)
nível de otimização total
descrição [42](#)
nível de rastreamento de erros
gerenciamento de logs de Serviços da Web [86](#)

O

objeto de dados personalizado
otimização de origens [22](#)
objetos de dados lógicos
armazenamento em cache no banco de dados [66](#)
origem de arquivo simples
otimização de origens [18](#)
otimização
efeitos colaterais [36](#)
método de otimização baseada em custos [45](#)
método de otimização dataship-join [46](#)
método de otimização de ajuste de ramificação [48](#)
método de otimização de empilhamento [49](#)
método de otimização de envio [49](#)
método de otimização de projeção antecipada [43](#)
método de otimização de seleção antecipada [48](#)
método de otimização de semi-associação [46](#)
métodos de desempenho do mapeamento [42](#)
otimização baseada em custos
descrição [45](#)
otimização com seleção antecipada
descrição [48](#)
Transformação de Consumidor de Serviço da Web [38](#)
transformação SQL [34](#)
otimização dataship-join
descrição [46](#)

otimização de ajuste de ramificação
descrição [48](#)
otimização de conexões
pool de conexões [87](#)
tamanho dos pacotes de rede do banco de dados [89](#)
otimização de consultas
otimização de origens [18](#)
otimização de conversão de tipos de dados
otimização de mapeamentos [52](#)
otimização de destinos
cargas em massa [15](#)
destino de arquivo simples [14](#)
intervalos de pontos de verificação do banco de dados [15](#)
otimização do banco de dados Oracle [16](#)
otimização de empilhamento
descrição [49](#)
otimização de envio
descrição [49](#)
habilitando na transformação SQL [35](#)
Transformação de Consumidor de Serviço da Web [38](#)
transformação SQL [34](#)
otimização de expressões
otimização de mapeamentos [25](#)
otimização de filtros
otimização de mapeamentos [52](#)
otimização de mapeamentos
leitura de passagem única [51](#)
otimização de conversão de tipos de dados [52](#)
otimização de expressões [25](#)
otimização de filtros [52](#)
rastreamento de erros [53](#)
otimização de origens
filtros condicionais [19](#)
objeto de dados personalizado [22](#)
origem de arquivo simples [18](#)
otimização de consultas [18](#)
otimização do banco de dados Oracle [23](#)
restrições [21](#)
selecionar distintas [19](#)
otimização de projeção antecipada
descrição [43](#)
Otimização de semiassociação
descrição [46](#)
otimização de Serviço de Dados SQL
drivers JDBC [71](#)
otimização de Serviços da Web
compactação de mensagens de Serviços da Web [80](#)
otimizar solicitações HTTP [80](#)
otimização de Serviços de Dados SQL
ferramentas de cliente de terceiros [71](#)
otimização de tempo de execução
estatísticas de monitoramento [63](#)
otimização do Serviço Analyst [61](#)
otimização do Serviço de Integração de Dados [61](#)
otimização do Serviço de Repositório do Modelo [62](#)
otimização do sistema [68](#)
otimização de transformações
cache de transformações [35](#)
eliminação de erros de transformação [36](#)
transformação de agregador [24](#)
Transformação de Associador [30](#)
transformação de Classificador [33](#)
Transformação de Consumidor de Serviço da Web [37](#)
transformação de Pesquisa [30](#)
transformação Java [27](#)
transformação SQL [34](#)
otimização do banco de dados Oracle
otimização de destinos [16](#)

- otimização do banco de dados Oracle ()
 - otimização de origens [23](#)
- otimização do Serviço Analyst
 - otimização de tempo de execução [61](#)
- otimização do Serviço de Integração de Dados
 - otimização de tempo de execução [61](#)
- otimização do Serviço de Repositório do Modelo
 - otimização de tempo de execução [62](#)
- otimização do sistema
 - otimização de tempo de execução [68](#)
- otimizar solicitações HTTP
 - otimização de Serviços da Web [80](#)

P

- paralelismo máximo
 - aumentando [55](#)
- particionamento
 - otimizando [54](#)
 - otimizando bancos de dados de destino [57](#)
 - otimizando bancos de dados de origem [57](#)
 - otimizando destinos de arquivo simples [56](#)
 - otimizando origens de arquivos simples [56](#)
 - otimizando transformações [58](#)
 - várias CPUs [55](#)
- planos de consultas SQL
 - exibindo [74](#)
- pool de conexões
 - otimização de conexões [87](#)
 - propriedades [88](#)
- porta de filtro
 - Transformação de Consumidor de Serviço da Web [38](#)
- propriedades do cache de conjunto de resultados
 - otimização de tempo de execução [77](#)

R

- rastreamento de erros
 - otimização de mapeamentos [53](#)
- restrições
 - configurando restrições [22](#)
 - otimização de origens [21](#)

S

- selecionar distintas
 - otimização de origens [19](#)
- Serviço da Web
 - alocação de memória [85](#)
 - solicitações simultâneas [82](#)
- Serviço de Dados SQL
 - alocação de memória [75](#)
- sistema
 - afunilamentos no UNIX, identificando [12](#)

- sistema ()
 - afunilamentos no Windows, identificando [11](#)
- solicitações simultâneas
 - Serviço da Web [82](#)
 - Serviço de Dados SQL [75](#)

T

- tabelas temporárias
 - descrição [78](#)
- tabelas virtuais
 - armazenamento em cache no banco de dados [66](#)
- tamanho dos pacotes de rede do banco de dados
 - otimização de conexões [89](#)
- Tem Efeitos Colaterais
 - descrição de propriedades de transformação [36](#)
- transformação de agregador
 - otimização de transformações [24](#)
- Transformação de Associador
 - otimização de transformações [30](#)
- transformação de Classificador
 - otimização de transformações [33](#)
- Transformação de Consumidor de Serviço da Web
 - habilitando a otimização de envio [39](#)
 - otimização com seleção antecipada [38](#)
 - otimização de envio [38](#)
 - otimização de filtros [38](#)
 - otimização de transformações [37](#)
- transformação de Pesquisa
 - otimização de transformações [30](#)
- transformação Java
 - otimização de transformações [27](#)
- transformação SQL
 - otimização com seleção antecipada [34](#)
 - otimização de envio [34](#)
 - otimização de transformações [34](#)
 - propriedades da otimização de envio [35](#)
- transformações
 - otimizando para particionamento [58](#)

U

- UNIX
 - afunilamentos do sistema [12](#)

W

- Windows
 - afunilamentos [11](#)