Utilizando o Change Data Capture no SQL Server 2008



Pedro Antonio Galvão Junior

pedrogalvaojunior@gmail.com

Atua no mercado de TI/SI há 14 anos, graduado no curso superior de Tecnologia em Sistemas da Informação Faculdade Uninove de São Roque, Pós-Graduado (Gestão e Engenharia de Processos para Desenvolvimento de Software com RUP) FIAP, formação MCDBA (SQL Server 2000), MVP 2007, 2008, 2009 e 2010 (Most Valuable Profissional) Windows Server System – SQL Server. Trabalha atualmente como Engenheiro de Processos e Administrador de Bancos de Dados SQL Server 2000 e 2005, na Fábrica de Artefatos de Látex São Roque. Professor Universitário nas Faculdades Academia de Ensino Superior de Sorocaba e São Roque, palestrante de Eventos TechEd 200, 2008 e 2009. Moderador dos fóruns MSDN Brasil e TechNet Brasil.

http://juniorgalvao-mvp2007.spaces.live.com

Com o passar do tempo o volume de informações armazenadas em uma tabela pode crescer de forma considerável, tornando-se um grande repositório. Este crescimento é provocado pela manipulação de dados, mais precisamente por inserções e atualizações que ocorrem constantemente em uma tabela.

Com o objetivo de analisar, identificar e documentar todas as ações que ocorrem em um banco de dados e suas tabelas, o SQL Server 2008 apresenta uma nova funcionalidade, chamada Change Data Capture ou Captura de dados de alterações.

Neste artigo apresentaremos esta nova funcionalidade, fornecida a partir das versões Enterprise, Developer e Evaluation.

Entendendo o Change Data Capture

O Change Data Capture é considerado uma das maiores inovações adicionadas ao SQL Server. Através desta funcionalidade torna-se possível realizar dois processos: o primeiro é chamado de captura dos dados alterados em tempo real e o segundo é denominado rastreamento de dados alterados.

O processo de captura de dados baseia-se na utilização dos comandos de manipulação de dados (INSERT, UPDATE e DELETE), mais conhecidos como comandos DML (Data Manipulation Language). Este processo utiliza um mecanismo de identificação e captura de dados chamado instância de captura de dados de alteração, conforme veremos nos próximos tópicos.

Esta instância possui a finalidade de observar a execução dos comandos DML e, conforme estes comandos são processados, a instância se encarrega de gerar uma cópia para serem posteriormente armazenadas em tabelas utilizadas pelo Change Data Capture.

O processso de rastreamento de dados alterados tem como finalidade pesquisar e informar quais os dados já foram alterados. Através de funções e stored procedures o SQL Server consegue obter estas informações.

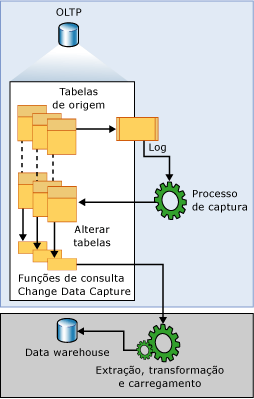
Com a utilização do Change Data Capture, o SQL Server cria um ambiente único para a realização destes dois processos, evitando a utilização de demais recursos ou técnicas complementares. Este ambiente é composto por diversos componentes específicos, como veremos aqui.

O funcionamento do Change Data Capture

O funcionamento do Change Data Capture pode ser considerado simples, tendo como base a execução de comandos DML. Mas existe outro elemento muito importante que disponibiliza as informações sobre os dados alterados. Este elemento é o Log de Transações ou Transaction Log, existente em qualquer banco de dados, responsável por armazenar informações sobre todos os procedimentos realizados.

A partir do log, o Change Data Capture identifica quais dados foram alterados, repassando estas informações para o processo de captura. Este, por sua vez, repassa este mesmo dado para as tabelas utilizadas para o armazenamento de dados alterados, ou seja, uma cópia do dado é armazenada em tabelas.

Essa sequência de atividades é conhecida como Fluxo de Dados Processados, como pode ser visto na **Figura 1**.



**Figura 1.** Fluxo de dados realizado durante o processo de captura de dados

Conforme os dados são inseridos de forma incremental em outras tabelas ou fontes de dados externas (data mart ou data warehouse), o Change Data Capture consegue capturar o volume de informações que está sendo manipulada a partir do log de transações.

Se considerarmos que este processo de captura poderá ser realizado constantemente em nosso ambiente durante um período de tempo, podemos criar alguns indicadores de informação que nos permitem dimensionar:

* Quantidade de linhas de registros processadas;
* Quantidade de inserções realizadas;
* Quantidade de atualizações realizadas;
* Volume ou tamanho de dados processados, etc.

Estes indicadores podem ser informações importantes para um administrador de banco de dados tomar diversas decisões relacionadas ao processamento e consumo de recursos realizadas sobre um banco de dados ou tabela.

O processo de captura se encerra a partir do momento que as funções de consulta de dados utilizadas pelo Change Data Capture são desabilitadas. Através do trabalho realizado pela instância de captura, torna-se possível catalogar e retornar informações sobre os dados alterados. Estas funções são conhecidas como CDC functions.

Além disso, o processo de captura de dados também pode ser aplicado em conjunto com ferramentas e tecnologias ETL, durante o processo de inserção de dados em tabelas ou fontes de dados externas, como data warehouse ou data marts.

Os recursos utilizados pelo Change Data Capture

Para realizar os processos de captura e rastreamento de dados alterados o SQL Server utiliza alguns recursos para trabalhar com o Change Data Capture. Organizados de acordo com sua funcionalidade, são eles:

* Tabela de Alteração;
* Instância de Captura de dados;
* Tabela de Origem;
* Change Data Capture Agent.

Tabela de Alteração

A tabela de alteração é associada a uma tabela que está sendo utilizada pelo Change Data Capture. Ela apresenta em sua estrutura algumas particularidades para possibilitar o controle do fluxo de dados ocorrido sobre determinada tabela. Dentre estas particularidades destacamos as cinco primeiras colunas de metadados, onde cada coluna fornece informações adicionais pertinentes às alterações registradas.

As colunas restantes espelham as colunas capturadas sobre a tabela de origem, respeitando o nome e tipo de dados de cada coluna. Cada operação de inserção ou exclusão que é aplicada a uma tabela de origem aparece como uma única linha dentro da tabela de alteração.

As colunas de dados da linha que são o resultado de uma operação de inserção contêm os valores de coluna depois da inserção. As colunas de dados da linha que são o resultado de uma operação de exclusão contêm valores de coluna antes da exclusão. Uma operação de atualização requer uma entrada de linha para identificar os valores da coluna antes da atualização e uma segunda entrada para identificar os valores da coluna depois da atualização.

Cada linha em uma tabela de alteração também contém metadados adicionais para permitir a interpretação da atividade de alteração.

* A coluna **\_\_$start\_lsn** identifica o número da seqüência do **log de confirmação** (**LSN**) que foi atribuído à alteração. O **LSN** de confirmação não só identifica alterações que foram confirmadas dentro da mesma transação, mas também ordena essas transações;
* A coluna **\_\_$seqval** pode ser usada para ordenar mais alterações que acontecem na mesma transação, por exemplo, uma sequência de registros que estão sendo inseridos em uma tabela. Através desta coluna podemos encontrar o número sequencial de inserções realizadas;
* A coluna **\_\_$operation** registra a operação associada à alteração. 1 = excluir, 2 = inserir, 3 = atualizar (indica a posição dos dados antes da alteração) e 4 = atualizar (indica a posição dos dados após alteração);
* A coluna **\_\_$update\_mask** é uma máscara de bit variável com um bit definido para cada coluna capturada. Para inserir e excluir entradas, a máscara de atualização sempre terá todos os bits definidos internamente pelo SQL Server. No entanto, as linhas de atualizações ocorridas sobre esta tabela serão definidas utilizando apenas um bit sobre a coluna ou colunas alteradas.

Instância de Captura de Dados

Antes do rastreamento das alterações que possam ocorrer em qualquer tabela, a captura de dados é habilitada para trabalhar sobre um banco de dados específico.

Isso representa para o SQL Server que determinado banco de dados terá a capacidade e recursos para realizar os processos de captura de dados. Para habilitar um banco de dados a utilizar o Change Data Capture é necessário utilizar a system stored procedure **sys.sp\_cdc\_enable\_db**.

Quando o banco de dados está habilitado é possível identificar as tabelas de origem controladas pelo Change Data Capture.

A instância de captura consiste em uma tabela de alteração e no máximo duas funções de consulta de dados, independente do tamanho do banco de dados e tabelas utilizadas no Change Data Capture.

Os metadados que descrevem os detalhes da configuração de uma instância de captura são mantidos nas tabelas de captura de dados de alteração **cdc.change\_tables**, **cdc.index\_columns** e **cdc.captured\_columns**. Estas informações podem ser recuperadas usando a system stored procedure **sys.sp\_cdc\_help\_change\_data\_capture**.

Todos os objetos associados a uma instância de captura são criados no esquema de captura de dados de alteração dentro do banco de dados habilitado para trabalhar com o CDC.

Os requisitos para o nome da instância de captura devem ser o mesmo válido para qualquer objeto, sendo este nome exclusivo entre as instâncias de captura de banco de dados.

Por padrão, o nome do objeto é composto por duas partes, **<nome do esquema de usuário>.<nome da tabela de origem>**, sendo, por exemplo, representado da seguinte forma:

* Nome do esquema de usuário = DBO.;
* Nome da tabela de origem = PRODUTOS;
* Formando o seguinte objeto utilizado pela instância de captura: DBO.PRODUTOS.

Após a configuração e habilitação da instância de captura de dados, o Change Data Capture acrescenta automaticamente a sigla de identificação **\_CT** ao nome da instância de captura, com base no nome da tabela associada a esta instância.

A função **fn\_cdc\_get\_all\_changes\_** é utilizada para consultar todas as alterações realizadas sobre determinada instância de captura. Como podemos utilizar diversas instâncias de captura, o SQL Server renomeia cada função **fn\_cdc\_get\_all\_changes\_** acrescentando ao nome desta função o nome de cada instância de captura de dados, conforme o exemplo **fn\_cdc\_get\_all\_changes\_dbo\_Produtos**.

**Tabela de Origem**

A definição da tabela de origem pode se tornar um pouco confusa, mas na verdade, trata-se de um termo criado para diferenciar as tabelas utilizadas pelo CDC no processo de carga em relação as tabelas criadas pelo usário, onde consideramos como tabela de origem, as tabelas criados pelo próprio usuário em seu respectivio banco de dados. Posteriormente esta mesma tabela, poderá ser configurada e habilitada para ser utilizada no Change Data Capture.

A tabela de origem possue os dados inseridos pelo usuário através de suas aplicações, processos de carga de dados ou manualmente, sendo estes, utilizados no processo de captura.

Utilizando a system stored procedure **sys.sp\_cdc\_enable\_table** podemos ativar o uso do CDC sobre uma determinada tabela de origem definida pela usuário

A captura de dados de alteração ajuda a reduzir o impacto gerado pelos consumidores (consultas de dados, relatório, etc.), permitindo que os conjuntos de resultados fornecidos permaneçam inalterados mesmo que a estrutura da coluna da tabela de origem seja alterada. Principalmente quando a tabela de origem estiver utilizando alguma coluna do tipo timestamp, sendo que, colunas timestamp não são utililzadas no processo de captura.

Neste caso, o processo de captura responsável pela população da tabela de alteração irá ignorar qualquer nova coluna não identificada durante o processo.

Através desta forma de trabalho, podemos garantir com o uso CDC a integridade e consistência dos dados que estão sendo capturados, sem que a instância CDC utilizada no processo de captura de dados, necessite de mais recursos durante seu processamento.

Caso uma coluna rastreada seja excluída ou removida do processo de captura, serão fornecidos valores nulos para a coluna nas entradas da tabela de alteração subseqüentes. Se uma coluna existente sofrer uma alteração em seu tipo de dados, a alteração será propagada para a tabela de alteração para garantir que o mecanismo de captura não perca qualquer dado manipulado.

O processo de captura também envia qualquer alteração detectada na estrutura da coluna de tabelas rastreadas para a tabela **cdc.ddl\_history**, e através da system stored procedure **sys.sp\_cdc\_get\_ddl\_history** torna-se possível gerar alertas sobre ajustes que devem ser feitos em aplicativos que possam utilizar estes dados.

Basicamente, a instância de captura atual continuará a manter sua forma quando alterações de DDL forem aplicadas à sua tabela de origem associada. Porém, é possível criar uma segunda instância de captura para a tabela que reflete a nova estrutura da coluna. Isso permite que o processo de captura consiga realizar simultâneamente as alterações em cada tabela de origem utilizada pelas instâncias CDC.

Portanto, uma tabela de alteração poderá ser utilizada para fornecer dados a seus respectivos programas em operação, sendo que, outra tabela conseguirá fornecer dados para outro ambiente, trabalhando distintamente sem qualquer impacto em ambas. Mesmo para tabelas que necessitem incorporar dados à nova coluna.

Isso permite ao mecanismo de captura que preencha ambas as tabelas de alteração, o que significa a execução de uma transação para cada tabela de forma simultânea, sem gerar qualquer tipo de perda de dados, mas proporcionando um pequeno aumento no tempo de processamento, para atender todas as transações configuradas para cada tabela. .

Change Data Capture Agent

Normalmente as atividades realizadas pelo SQL Server Agent estão associadas a um banco de dados ou demais objetos existentes em uma instância SQL Server. Além disso, a partir do momento em que o Change Data Capture é habilitado, o SQL Server Agent sofrerá algumas alterações em seu conjunto de trabalhos (Jobs).

Com base nesta alteração são adicionados dois novos trabalhos (Jobs) associados ao Change Data Capture, para isso é necessário que o SQL Server Agent esteja ativo.

O primeiro trabalho é utilizado para preencher as tabelas de alteração do banco de dados e o outro é responsável pela limpeza da tabela de alteração. Ambas atividades consistem em uma única etapa que executa um comando Transact-SQL. Este comando representa uma system stored procedure utilizada na captura de dados de alteração, com toda a estrutura necessária para execução dos trabalhos configurados no SQL Server Agent.

As atividades executadas pelo Change Data Capture Agent são criadas quando a primeira tabela do banco de dados é habilitada para captura de dados de alterações. O trabalho de limpeza é criado automaticamente, já o trabalho de captura só será criado se não existir nenhum outro trabalho de captura com o mesmo nome neste banco de dados.

O trabalho de captura também poderá ser criado quando a captura de dados de alteração estiver habilitada para um banco de dados. Além disso, o trabalho de leitura do log de transações definido para esta instância deverá ter acesso às transações de alterações de dados.

Tanto os trabalhos de captura como os de limpeza são criados com base nos parâmetros padrões definidos pelo SQL Server. O trabalho de captura é iniciado imediatamente e executado continuamente, processando um máximo de 1000 transações por ciclo de verificação com uma espera de cinco segundos entre os ciclos. O trabalho de limpeza é executado diariamente a cada duas horas, ele retém as entradas da tabela de alteração por três dias, removendo um máximo de 5000 entradas com uma única instrução de exclusão.

Os trabalhos do agente de captura de dados são removidos quando o Change Data Capture é desabilitado para um banco de dados, ou até a primeira alteração de dados ocorrer.

Internamente, os trabalhos de agente do Change Data Capture são criados e descartados usando as systems stored procedures **sys.sp\_cdc\_add\_job** e **sys.sp\_cdc\_drop\_job**, respectivamente.

O administrador do banco de dados não tem nenhum controle explícito sobre a configuração padrão das atividades realizadas pelo Change Data Capture Agent. Para isso, a system stored procedure **sys.sp\_cdc\_change\_job** é fornecida para permitir que os parâmetros de configuração padrão sejam modificados. Além disso, a system stored procedure **sys.sp\_cdc\_help\_jobs** permite exibir os parâmetros de configuração atual.

Os trabalhos de captura e de limpeza extraem os parâmetros de configuração da tabela **msdb.dbo.cdc\_jobs** na inicialização. Qualquer alteração feita nesses valores utilizando a system stored procedure **sys.sp\_cdc\_change\_job** não entrará em vigor até que o trabalho seja interrompido e reiniciado.

Duas systems stored procedures são adicionadas para permitir que os trabalhos de agente do Change Data Capture sejam iniciados e interrompidos: **sys.sp\_cdc\_start\_job** e **sys.sp\_cdc\_stop\_job**.

É importante destacar que iniciar e interromper o trabalho de captura não resulta em perda de dados sobre o processo de captura que está em execução. Este procedimento impede que o processo de captura verifique ativamente o log de transações, ou consulte as informações armazenadas nas tabelas de alteração utilizadas pelo Change Data Capture.

Uma estratégia para impedir que a verificação do log adicione carga durante os períodos de pico de demanda no acesso ao banco de dados é interromper o trabalho de captura e reiniciá-lo quando houver redução de demanda.

Os dois trabalhos do Change Data Capture Agent foram projetados para serem flexíveis e configuráveis o suficiente para atender às necessidades básicas dos ambientes do Change Data Capture.

Componentes utilizados pelo Change Data Capture

Para realizar os processos de captura e rastreamento de dados alterados o SQL Server cria alguns componentes específicos para trabalhar com o Change Data Capture, organizados de acordo com sua funcionalidade:

* Tabelas do Change Data Capture;
* Stored Procedures do Change Data Capture;
* Functions do Change Data Capture.

Tabelas do Change Data Capture

Com base nos comandos de manipulação de dados o Change Data Capture registra as informações sobre as capturas de dados realizadas, armazenando estas e outras informações em tabelas específicas.

Outra característica apresentada pelo Change Data Capture é a capacidade de interagir com os comandos DDL, onde cada alteração sofrida na estrutura de uma tabela habilitada também pode ser monitorada pela instância de captura de dados vinculada a esta tabela.

Para todos os processos de rastreamento, monitoramento e controle de dados, o Change Data Capture utiliza tabelas específicas criadas pelo SQL Server durante seu processo de configuração. A seguir você encontrará informações sobre as tabelas utilizadas pelo Change Data Capture:

* **cdc.<capture\_instance>\_CT:** Ao consultar esta tabela será retornada uma linha para cada alteração feita em uma coluna capturada na tabela de origem associada;
* **cdc.captured\_columns:** Ao consultar esta tabela será retornada uma linha para cada coluna rastreada em uma instância de captura;
* **cdc.change\_tables:** Ao consultar esta tabela será retornada uma linha para cada tabela de alteração do banco de dados;
* **cdc.ddl\_history:** Ao consultar esta tabela será retornada uma linha para cada alteração de DDL feita nas tabelas que estão habilitadas para o Change Data Capture;
* **cdc.lsn\_time\_mapping:** Ao consultar esta tabela será retornada uma linha para cada transação que tem linhas em uma tabela de alteração. Esta tabela é usada para mapear os valores confirmados de LSN e a hora em que a transação foi confirmada;
* **cdc.index\_columns:** Ao consultar esta tabela será retornada uma linha para cada coluna de índice associada a uma tabela de alteração;
* **dbo.cdc\_jobs (Transact-SQL):** Ao consultar esta tabela será retornado os parâmetros de configuração para trabalhos do agente Change Data Capture.

Stored Procedures do Change Data Capture

O Change Data Capture disponibiliza em formato relacional o histórico de atividades DML ocorrido em cada tabela. Este formato relacional consiste na utilização dos respectivos relacionamentos entre as tabelas de origem e as tabelas utilizadas pelo Change Data Capture durante o processo de captura dos dados.

As stored procedures apresentadas na **Tabela 1** são consideradas as mais utilizadas para o processo de configuração o Change Data Capture. Por exemplo, através da system stored procedure **sys.sp\_cdc\_start\_job** podemos iniciar um trabalho (Job) configurado no SQL Server Agent para realizar alguma atividade necessária para o Change Data Capture.

|  |  |
| --- | --- |
| **Systems Stored Procedure** | **Descrição** |
| sys.sp\_cdc\_disable\_db | Desabilita a captura de dados de alteração do banco de dados atual. |
| sys.sp\_cdc\_disable\_table | Desabilita a captura de dados de alteração para a tabela de origem especificada e a instância de captura no banco de dados atual. |
| sys.sp\_cdc\_enable\_db | Habilita o Change Data Capture para o banco de dados atual. |
| sys.sp\_cdc\_enable\_table | Habilita o Change Data Capture para a tabela de origem especificada no banco de dados atual. Quando uma tabela está habilitada para Change Data Capture, um registro de cada operação DML (Linguagem de Manipulação de Dados) aplicado à tabela é gravado no log de transações. |
| sys.sp\_cdc\_help\_change\_data\_capture | Retorna a configuração de captura dos dados de alteração para cada tabela habilitada para a captura de dados de alteração no banco de dados atual. Podem ser retornadas até duas linhas para cada tabela de origem, uma linha para cada instância de captura. |
| sys.sp\_cdc\_scan | Executa a operação de verificação no log Change Data Capture. |

**Tabela 1.** Relação de systems stored procedures mais utilizadas.

Para saber mais sobre outras tabelas existentes para se trabalhar com o Change Data Capture, acesse o Books On-Line do SQL Server 2008.

Functions do Change Data Capture

Através do registro de atividades ocorridas sobre cada tabela, algumas functions são utilizadas para fornecer informações ou detalhes de cada alteração sofrida por estas tabelas. Estas funções facilitam a consulta e o consumo destes dados de forma simples e prática.

O processo de consulta de dados disponibilizado para o Change Data Capture consiste na atividade de espelhar os dados entre a tabela de origem e tabela de captura. Assim, as functions utilizadas pelo Change Data Capture conseguem consultar estas tabelas de alteração apresentando os dados que foram manipulados, dentre outras informações.

Para retornar informações sobre cada manipulação de dados, ou outras informações específicas do Change Data Capture, podemos utilizar diversas system functions, conforme apresenta a **Tabela 2**. Para saber mais sobre estas system functions acesse o Books On-Line do SQL Server 2008.

|  |  |
| --- | --- |
| **Systems Functions** | **Descrição** |
| cdc.fn\_cdc\_get\_all\_changes\_<capture\_instance> | Retorna uma linha para cada alteração aplicada à tabela de origem dentro do intervalo LSN (número de seqüência de log) especificado. Se uma linha de origem tiver passado por várias alterações durante o intervalo, todas as alterações serão representadas no conjunto de resultados retornado. Além de retornar os dados de alteração, quatro colunas de metadados fornecem as informações necessárias para a aplicação de alterações em outra fonte de dados. |
| cdc.fn\_cdc\_get\_net\_changes\_<capture\_instance> | Retorna uma linha de alteração líquida para cada linha de origem alterada dentro do intervalo LSN especificado. Ou seja, quando uma linha de origem tiver várias alterações durante o intervalo de LSN, uma única linha que reflete o conteúdo final da linha será retornada pela função. |
| sys.fn\_cdc\_get\_min\_lsn | Retorna o valor da coluna start\_lsn para a instância de captura especificada na tabela do sistema cdc.change\_tables. Esse valor representa o ponto de extremidade inferior do intervalo de validade da instância de captura. |
| sys.fn\_cdc\_get\_max\_lsn | Retorna o LSN (número de seqüência de log) máximo da coluna start\_lsn na tabela do sistema cdc.lsn\_time\_mapping. É possível usar essa função para retornar o ponto de extremidade superior da linha do tempo de captura de dados de alteração para qualquer instância de captura. |
| sys.fn\_cdc\_has\_column\_changed | Identifica se a máscara de atualização especificada indica que a coluna especificada foi atualizada na linha de alteração associada. |

**Tabela 2.** Relação de systems functions mais utilizadas.

Configurando o Change Data Capture

Com o objetivo de garantir um modelo seguro, durante o processo de configuração do CDC o SQL Server cria um usuário especial e um esquema de captura de dados de alteração sobre o banco de dados selecionado.

Os objetos utilizados pelo Change Data Capture são configurados sobre este esquema e pertecem exclusivamente ao usuário definido na configuração para trabalhar com CDC. Os demais objetos existentes sobre este banco de dados não sofrem nenhum tipo de alteração.

Além disso, a partir do momento em que o Change Data Capture é configurado, algumas systems stored procedures são criadas para fornecer suporte aos trabalhos de captura de dados de alteração, pertencendo de forma restrita a membros da **server role** **sysadmin** e membros da **database role db\_owner**.

Habilitando o Change Data Capture para um banco de dados

Para habilitar ou desabilitar o Change Data Capture sobre o contexto do banco de dados, podemos utilizar a system stored procedure **sys.sp\_cdc\_enable\_db** ou **sys.sp\_cdc\_disable\_db**, respectivamente, necessitando ser um membro da **server role sysadmin**.

Vamos iniciar a demonstração de uso e habilitação do Change Data Capture criando um novo banco de dados chamado SQLMagazine, conforme apresentado na **Listagem 1**. Em seguida verificamos o status do Change Data Capture para este banco consultando a system view catalog **sys.databases** (**Figura 2**).

**Listagem 1.** Criação do Banco de dados

-- Bloco 1 --

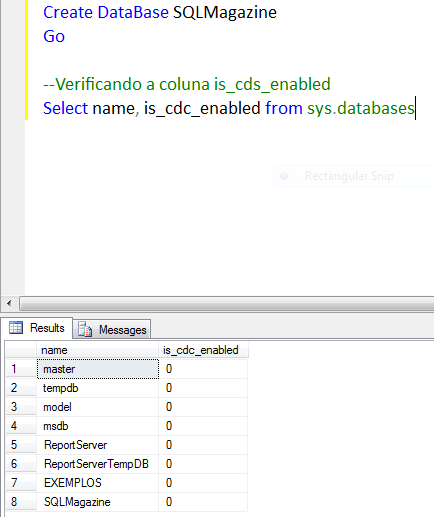
Create Database SQLMagazine

Go

-- Bloco 2 --

Select Name, Is\_cdc\_Enabled from Sys.Databases

Go



**Figura 2.** Relação de Bancos de Dados existentes na instância SQL Server.

Após verificar o status da configuração do Change Data Capture, podemos notar que o nosso banco de dados ainda não está habilitado e configurado para trabalhar com esta funcionalidade. Para isso, utilizamos a system stored procedure **sys.sp\_cdc\_enable\_db**, de acordo com a **Listagem 2**. Mais uma vez é importante ressaltar que para executar este procedimento é necessário que o usuário seja membro da **server role sysadmin.**

**Listagem 2.** Habilitando o change data captura para o banco de dados

-- Bloco 1 --

Use SQLMagazine

Go

-- Bloco 2 --

Execute sys.sp\_cdc\_enable\_db;

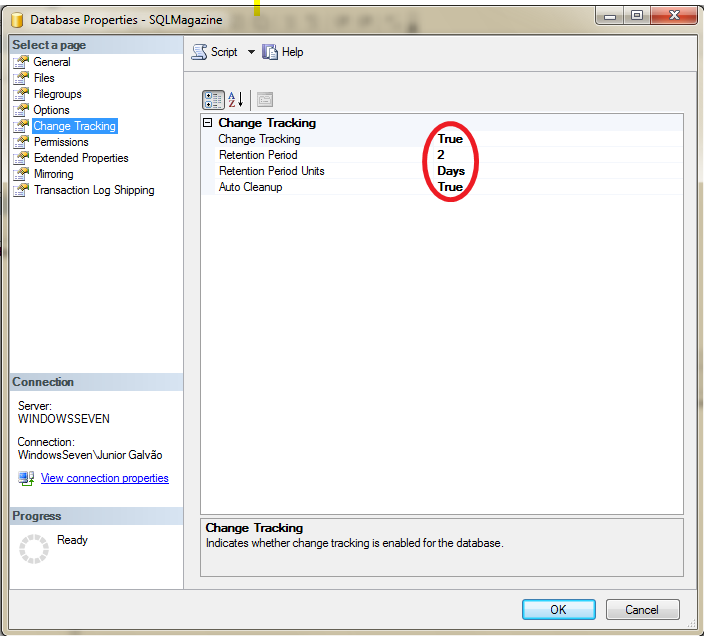
Go

Após executar **sys.sp\_cdc\_enable\_db** o SQL Server criará todos os objetos do Change Data Capture no banco de dados SQLMagazine, inclusive tabelas de metadados e gatilhos **DDL**.

O SQL Server também cria o esquema **cdc** e **usuário de banco de dados cdc**, atribuindo para a coluna **is\_cdc\_enabled** existente no banco de dados na system view catalog (exibição do catálogo) **sys.databases**, o valor 1, que representa que este banco de dados esta configurado e habilitado para utilizar o Change Data Capture.

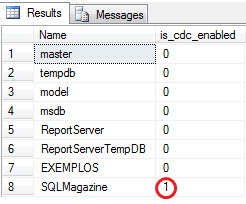
Caso um esquema ou usuário denominado **cdc** existir no banco de dados, o mesmo não poderá ser habilitado para Change Data Capture até que o esquema e/ou o usuário sejam removidos ou renomeados.

A seguir você poderá observar as propriedades do Banco de Dados SQLMagazine após a execução da **sys.sp\_cdc\_enable\_db**, confome apresenta a **Figura 3.**

****

**Figura 3.** Propriedade do Banco de Dados SQLMagazine.

Vamos agora confirmar o status do Change Data Capture após sua habilitação para o banco SQLMagazine. Para isso, devemos observar a coluna **is\_cdc\_enabled**, que deve apresentar o valor **1**, na view catalog **sys.databases** (**Figura 4**).



**Figura 4.** Banco SQLMagazine habilitado para o CDC.

Desabilitando o Change Data Capture para um banco de dados

Ao desabilitar a utilização do Change Data Capture o SQL Server realiza todo processo inverso de configuração do banco de dados. Para realizar este procedimento é necessário que o usuário seja membro da **server role sysadmin**. Com isso, o usuário poderá executar a system stored procedure **sys.sp\_cdc\_disable\_db**, como apresenta a **Listagem 3**.

Durante o processo de desativação do CDC todas as tabelas, usuários, esquemas e trabalhos configuradas para trabalhar com ele serão automaticamente desabilitadas.

No entanto, qualquer função associada que foi criada pelo Change Data Capture não será removida automaticamente, deverá ser excluída explicitamente.

Se um banco de dados habilitado para o processo de captura for excluído, todos os jobs utilizados pelo SQL Server Agent, associados ao CDC serão removidos automaticamente.

**Listagem 3.** Desabilitando o change data capture para banco de dados

-- Bloco 1 --

Use SQLMagazine

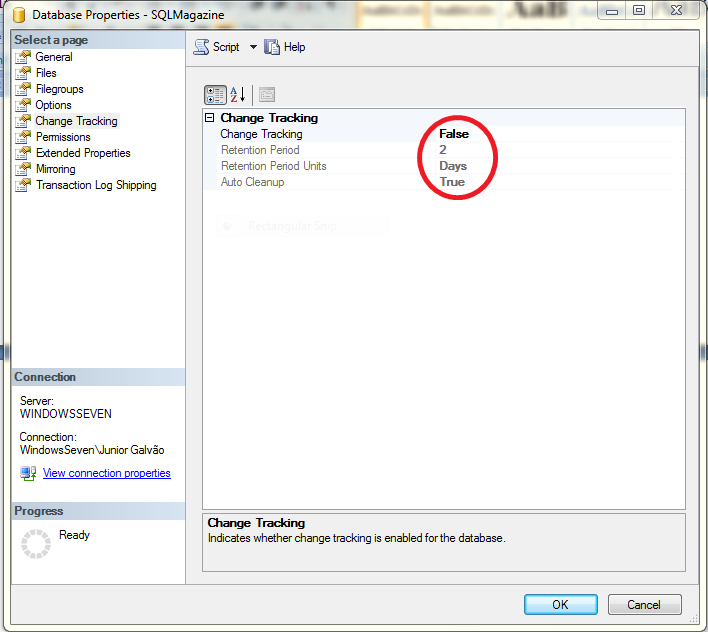
Go

-- Bloco 2 --

Execute sys.sp\_cdc\_disable\_db;

Go

A **sys.sp\_cdc\_disable\_db** remove todos os objetos relacionados ao Change Data Capture que têm escopo em todo o banco de dados, inclusive tabelas de metadados e gatilhos **DDL**. A **Figura 5** apresenta as propriedades do Banco de Dados SQLMagazine após a desativação do CDC.

****

**Figura 5.** Propriedade do Banco de Dados SQLMagazine.

Habilitando o Change Data Capture para uma tabela

Para habilitar ou desabilitar o Change Data Capture no nível de tabela devemos utilizar a system stored procedure **sys.sp\_cdc\_enable\_table** ou **sys.sp\_cdc\_disable\_table**. Para isso é necessário ser membro da **server role** **sysadmin** ou **membro da database role db\_owner**.

Vamos iniciar a demonstração de uso e habilitação do Change Data Capture sobre tabelas criando uma tabela chamada **Produtos** dentro do banco **SQLMagazine** (**Listagem 4**).Logo após, preparamos o ambiente configurando uma chave primária para esta tabela e a populando.

**Listagem 4.** Criando a Tabela Produtos e preparando o ambiente

-- Bloco 1 --

Use SQLMagazine

Go

-- Bloco 2 --

Create Table Produtos

(Codigo Int Identity(1,1),

Descricao VarChar(20))

Go

-- Bloco 3 – Adicionando a Chave Primária --

Alter Table Produtos

Add Constraint [PK\_Codigo\_Produtos] Primary Key (Codigo)

Go

-- Bloco 4 - Inserindo a Massa de Registros para Teste --

Declare @ContadorRegistros Int

Set @ContadorRegistros=1

While @ContadorRegistros <=1000

Begin

If @ContadorRegistros =1

Insert Into Produtos Values ('Produto Nº: 1')

Else

Insert Into Produtos Values ('Produto Nº: '+Convert(VarChar(4),@@Identity+1))

Set @ContadorRegistros += 1;

End

Go

Após a criação da tabela Produtos, definição da chave primária e população da tabela, o ambiente estará devidamente preparado para ser utilizado pelo CDC no processo de captura de informações.

Vamos dar início ao processo de configuração da captura de informações, utilizando tabela Produtos. Para inicara configuração, será utilizado a system stored procedure **sp\_cdc\_enable\_table**, responsável em habilitar a tabela Produtos para ser utilizado pelo Change Data Capture, conforme apresenta (**Listagem 5**).

**Listagem 5.** Habilitando o CDC para a tabela Produtos

-- Bloco 1 --

Use SQLMagazine

Go

-- Bloco 2 --

EXECUTE sys.sp\_cdc\_enable\_table

@source\_schema = N'dbo',

@source\_name = N'Produtos',

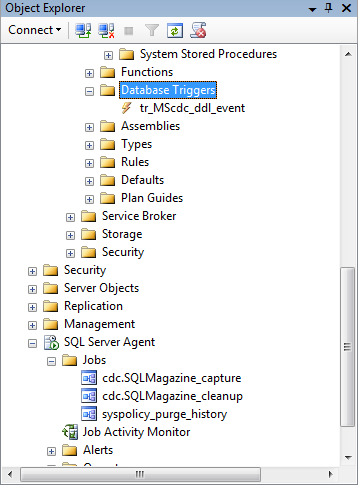
@role\_name = N'cdc\_Admin';

Go

Após a tabela ser habilitada, a instância de captura associada é criada para dar suporte à distribuição dos dados de alteração coletados em sua respectiva tabela de origem. Além disso, nossa instância SQL Server acaba de sofrer algumas mudanças em sua estrutura.

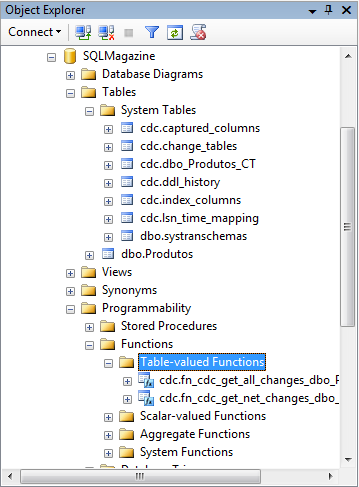
Neste momento o SQL Server acaba de adicionar e configurar sobre a instância SQL Server, mais especificamente sobre o SQL Server Agent, alguns jobs relacionados ao Change Data Capture.

Estes jobs serão utilizados para captura e limpeza da tabela Produtos, como podemos observar na **Figura 6**. Para que o processo de configuração destes jobs seja realizado com sucesso é necessário que serviço SQL Server Agent esteja iniciado.



**Figura 6.** Relação de jobs criados e configurados para trabalhar com o CDC

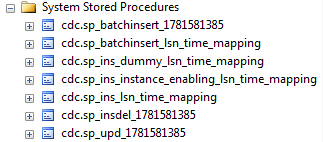
Agora que os jobs estão criados e configurados o SQL Server continuará a realizar alterações em nosso ambiente, neste caso, no banco de dados **SQLMagazine**, criando novas **systems tables** e **systems functions**, conforme a **Figura 7**.

****

**Figura 7.** Systems tables e systems functions criadas e configurados para trabalhar com o CDC.

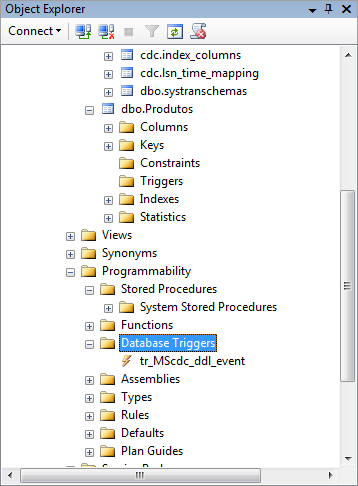
Podemos observar que o SQL Server adicionou ao nosso banco de dados diversas system tables e system functions, utilizadas pelo Change Data Capture, nos trabalhos de captura e limpeza.

Em seguida o processo de configuração do Change Data Capture (com base na tabela Produtos) continua com a criação de algumas system stored procedures, como na **Figura 8**, dentro do banco de dados SQLMagazine.

****

**Figura 8.** Relação de systems procedures criadas e configurados para trabalhar com o Change Data Capture.

O processo de configuração do Change Data Capture se encerra com a criação de database triggers, responsáveis por controlar as ações realizados pelo Change Data Capture durante a execução dos trabalhos de captura (**Figura 9**).

****

**Figura 9.** Database triggers criadas e configuradas para trabalhar com o CDC.

Demonstrando a captura de dados

Agora que estamos com todo o ambiente criado, configurado e liberado para ser utilizado sobre a tabela **Produtos**, vamos realizar alguns testes e observar todo controle e captura de dados realizado pelo SQL Server.

Durante a execução do trabalho de captura todos os dados coletados são armazenados dentro da **instância de Captura de Dados de Alteração**.Estes dados podem ser consultados e identificados através da **system function cdc.fn\_cdc\_get\_all\_changes\_dbo\_Produtos**, criada após a configuração da tabela Produtos.

A **Listagem 6** apresenta um exemplo onde realizaremos a inserção de novos dados na tabela Produtos, para posteriormente realizar a consulta das linhas capturadas. No final poderemos observar as informaões apresentadas sobre os dados capturados, conforme os resultados apresentados na **Figura 10 e Figura 11.**

**Listagem 6.** Captura e retorno de dados controlados pelo Change Data Capture

-- Bloco 1 --

Use SQLMagazine

Go

-- Bloco 2 - Inserindo novos dados na tabela Produtos --

Insert Into Produtos Values ('Produto Nº: '+Convert(VarChar(4),@@Identity+1))

Go

Insert Into Produtos Values ('Produto Nº: '+Convert(VarChar(4),@@Identity+1))

Go

-- Bloco 3 - Retornando todas as linhas capturadas pelo Change Data Capture --

Declare @from\_lsn binary(10),

@to\_lsn binary(10)

Set @from\_lsn = sys.fn\_cdc\_get\_min\_lsn('dbo\_Produtos')

Set @to\_lsn = sys.fn\_cdc\_get\_max\_lsn()

Select \* From cdc.fn\_cdc\_get\_all\_changes\_dbo\_Produtos (@from\_lsn, @to\_lsn, N'all')

Go

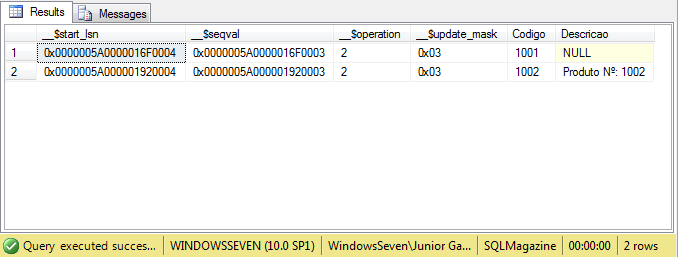
-- Bloco 4 – Atualizando dados existentes na tabela Produtos --

Update Produtos

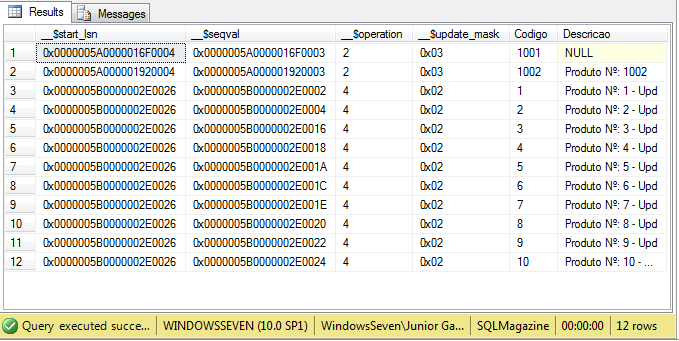
Set Descricao= Descricao+' - Upd'

Where Codigo Between 1 And 10

Go



**Figura 10.** Relação de linha de registros inseridas capturadas pelo Change Data Capture.



**Figura 11.** Relação de linha de registros atualizadas capturadas pelo Change Data Capture.

**Desabilitando o Change Data Capture para uma tabela**

Os membros da database role **db\_owner** ou da server role **sysadmin** poderão remover uma instância de captura para tabelas de origem individuais usando a system stored procedure **sys.sp\_cdc\_disable\_table** (veja a **Listagem 7**).

Para verificar se uma tabela de origem está habilitada examine a coluna **is\_tracked\_by\_cdc** na system view catalog **sys.tables**, observando se o valor retornado por esta coluna é **1**. Se uma tabela habilitada for descartada, os metadados associados a esta tabela (utilizados pelo CDC) também serão removidos automaticamente, e, se não restarem tabelas habilitadas para o banco de dados, os jobs também serão removidos.

**Listagem 7.** Desabilitando o Change Data Capture para tabela Produtos

-- Bloco 1 --

Use SQLMagazine

Go

-- Bloco 2 --

EXECUTE sys.sp\_cdc\_disable\_table

@source\_schema = N'dbo',

@source\_name = N'Produtos',

@capture\_instance = N'dbo\_Produtos;

Go

Podemos observar após a remoção do CDC para a tabela Produtos que todos os objetos (Stored Procedures, Functions e Tables) foram removidos, incluindos os jobs para captura e limpeza.

**Administrando e Gerenciando o Change Data Capture**

Para finalizar a demonstração, apresentaremos duas das mais importantes systems functions de administração e gerenciamento de uma instância Change Data Capture.

A primeira system function é conhecida como **sys.sp\_cdc\_get\_captured\_columns**.Esta function retorna informações sobre os dados coletados em uma tabela de origem utilizada nos trabalhos de captura. Além disso, apresenta informações sobre cada coluna da tabela de origem capturada.

A **Tabela 3** apresenta informações sobre cada coluna retornada pela **sys.sp\_cdc\_get\_captured\_columns**.

|  |  |  |
| --- | --- | --- |
| **Nome da coluna** | **Tipo de dados** | **Descrição** |
| source\_schema | **sysname** | Nome do esquema da tabela de origem. |
| source\_table | **sysname** | Nome da tabela de origem. |
| capture\_instance | **sysname** | Nome da instância de captura. |
| column\_name | **sysname** | Nome da coluna de origem capturada. |
| column\_id | **int** | ID da coluna na tabela de origem. |
| ordinal\_position | **int** | Posição da coluna na tabela de origem. |
| data\_type | **sysname** | Tipo de dados da coluna. |
| character\_maximum\_length | **int** | Comprimento máximo de caracteres que podem ser armazenados na coluna. Somente serão considerados os valores caracteres, caso contrário o SQL Server retorna o valor NULL. |
| NUMERIC\_PRECISION | **tinyint** | Precisão da coluna com base numérica. Se a precisão numérica não for compatível com valores tinyint, o SQL Server retorna o valor NUL . |
| numeric\_precision\_radix | **smallint** | Precisão da coluna com base numérica. Se a precisão numérica não for compatível com valores smallint, o SQL Server retorna o valor NULL |
| numeric\_scale | **int** | Escala da coluna com base numérica. Se a precisão numérica não for compatível com valores Int, o SQL Server retornar o NULL. |
| datetime\_precision | **smallint** | Precisão da coluna com base em data e hora. Se o valor informado não for reconhecido como um valor de data e hora o SQL Server retorna NULL. |

**Tabela 3.** Relação de colunas retornadas pela **sys.sp\_cdc\_get\_captured\_columns.**

A **Listagem 8** apresenta um exemplo de como utilizar **sys.sp\_cdc\_get\_captured\_columns** para retornar informações sobre as colunas que estão sendo capturadas para a tabela Produtos, conforme apresenta a (**Figura 12**).

**Listagem 8.** Retorna as colunas capturadas pelo Change Data Capture

-- Bloco 1 --

Use SQLMagazine

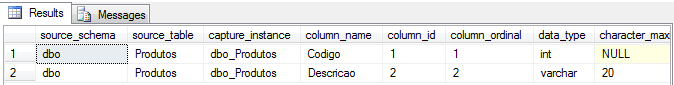
Go

-- Bloco 2 --

Execute sys.sp\_cdc\_get\_captured\_columns

@capture\_instance = N'dbo\_Produtos';

Go



**Figura 12.** Relação das colunas capturadas pelo Change Data Captura com base na tabela produtos.

Conforme apresentou a **Figura 12**, a **sys.sp\_cdc\_get\_captured\_columns** retornou uma linha para cada coluna existente na tabela Produtos, apresentando informações sobre estas respectivas colunas capturadas pelo CDC.

Outra system function bastante utilizada para gerenciamento e administração dos dados ou tabela capturadas é conhecida como **sys.sp\_cdc\_help\_change\_data\_capture**. Ela retorna a configuração do Change Data Capture com base em cada tabela utilizada na captura de dados de alteração. A **Tabela 4** apresenta as informações sobre cada coluna retornada por **sys.sp\_cdc\_help\_change\_data\_capture**.

|  |  |  |
| --- | --- | --- |
| **Nome da coluna** | **Tipo de dados** | **Descrição** |
| source\_schema | **sysname** | Nome do esquema da tabela de origem. |
| source\_table | **sysname** | Nome da tabela de origem. |
| capture\_instance | **sysname** | Nome da instância de captura. |
| object\_id | **int** | ID da tabela de alteração associada à tabela de origem. |
| source\_object\_id | **int** | ID da tabela de origem. |
| start\_lsn | **binary(10)** | LSN (Número de Sequência de Log) representando o ponto de extremidade inferior para consulta da tabela de alteração.  NULL = Nenhum ponto de extremidade inferior foi definido o SQL Server. |
| end\_lsn | **binary(10)** | LSN representando o ponto de extremidade superior para consulta da tabela de alteração. |
| supports\_net\_changes | **bit** | Suporte a captura de dados de alteração executadas através da rede. |
| has\_drop\_pending | **bit** | Não é usado no SQL Server 2008. |
| role\_name | **sysname** | Nome da função do banco de dados usada para controlar o acesso aos dados de alteração, caso nenhuma função seja especificada ou identificado, será retornado NULL. |
| index\_name | **sysname** | Nome do índice usado para identificar exclusivamente linhas na tabela de origem. |
| filegroup\_name | **sysname** | Nome do grupo de arquivos no qual a tabela de alteração reside, se a tabela residir no Filegroup Primary, será retornado NULL. Se a tabela residir em outro filegroup, será exibido o nome do filegroup. |
| create\_date | **datetime** | Data em que a instância de captura foi habilitada. |
| index\_column\_list | **nvarchar(max)** | Lista de colunas de índice usadas para identificar exclusivamente linhas na tabela de origem. |
| captured\_column\_list | **nvarchar(max)** | Lista de colunas de origem capturadas. |

**Tabela 4.** Relação de colunas retornadas pela **sys.sp\_cdc\_help\_change\_data\_capture.**

A **Listagem 9** apresenta um exemplo de como utilizar **sys.sp\_cdc\_help\_change\_data\_capture** para retornoar a relação de tabelas utilizadas atualmente pelo CDC em seus trabalhos de captura. O resultado é apresentado na **Figura 13.**

**Listagem 9.** Retorna a relação de tabelas utilizadas pelo Change Data Capture

-- Bloco 1 --

Use SQLMagazine

Go

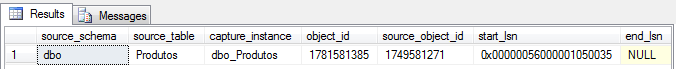
-- Bloco 2 --

Execute sys.sp\_cdc\_help\_change\_data\_capture

@source\_schema = N'dbo',

@source\_name = N'Produtos';

Go

****

**Figura 13.** Relação de tabelas controladas pelo Change Data Capture.

Neste exemplo podemos observar que somente uma linha foi retornada após utilizar a system function **sys.sp\_cdc\_help\_change\_data\_capture**, pois somente a tabela Produtos está configurada para trabalhar com o Change Data Capture.

Conclusão

Utilizar o Change Data Capture para uma instância de captura de dados representa um grande avanço para os administradores de bancos de dados e desenvolvedores.

Através do Change Data Capture temos a possibilidade que criar um histórico de manipulação de dados sem ter a necessidade de utilizar ferramentas de terceiros ou técnicas que possam gerar grandes impactos ou aumento no processamento durante a manipulação de dados no SQL Server 2008.

O artigo demonstrou o conceito e a prática deste recurso presente no SQL Server 2008 Enterprise Edition, Developer Edition e Evalution Edition. Aprendemos com os exemplos, como utilizar o Change Data Capture sobre bancos de dados e tabelas. Destacamos suas principais considerações e impactos em relação ao ambiente utilizado para se trabalhar com captura de dados de alteração.

***De que se trata o artigo:***

Neste artigo conheceremos as formas de captura de dados existente no SQL Server 2008, através do Change Data Capture. Em seguida, demonstraremos como realizar a captura destes dados, com base em uma tabelas contendo dados fictícios.

***Para que serve:***

Através do Change Data Capture, torna-se possível a criação de um repositório ou histórico de dados, oferecendo a sua organização criar e manter de base de conhecimento. Esta base de conhecimento, deverá ajudar a estabelecer o nível de importância sobre as informações manipuladas diariamente, fazendo com que os Diretores, Gerentes e Administradores avaliem o quanto esta informação é importante e como ela poderá ajudar na tomada de decisão á curto ou médio prazo.

***Em que situação o tema útil:***

A captura de dados é uma técnica útil para ambientes que apresentam processamento diário de informações. Através dos trabalhos de captura de dados, podemos criar uma organização sobre os dados manipulados, estabelecendo regras de controle e auditoria. Utilizando os recursos existentes no Change Data Capture, torna-se possível automatizar ainda mais as formas de coletada e categorização de informações.

Links

**WebCast - Semana do SQL Server 2008 : Data Warehousing - Escalabilidade e Performance**

http://msevents. .com/CUI/EventDetail.aspx?EventID=1032383692&culture=pt-br

**WebCast - Visão Geral do Gerenciamento de Dados do SQL Server 2008**

http://msevents. .com/CUI/EventDetail.aspx?EventID=1032354711&culture=pt-br

**Manuais Online do SQL Server 2008.**

http://msdn. .com/pt-br/sqlserver/cc514207.aspx

**SQL Server MSDN Experience.**

http://www.msdnbrasil.com.br/experience/sqlserver/