



عنوان مقاله: Change Data Capture یا CDC چیست؟

نویسنده مقاله: علیرضا ارومند

تاریخ انتشار: خرداد ۹۴

منبع: <https://nikamooz.com/change-data-capture>

مقدمه

یکی از قابلیت های ویژه SQL Server Enterprise Edition (که البته در نسخه های Developer و editions هم به خوبی کار می کند) است که روی جدولی در دیتابیس ما فعال می شود و تغییرات ناشی از دستورات (Insert, Update, Delete) DML آن را نگهداری می کند و ما میتوانیم از این اطلاعات استفاده کنیم. وقتی که ما CDC را روی جدولی فعال میکنیم، SQL Server جدولی مشابه آن به همراه یکسری Metadata ایجاد می کند و تغییرات را در آن ذخیره مینماید. در کنار آن تعدادی Table-valued function در اختیار ما قرار میگیرند که به ما امکان استفاده از داده های ذخیره شده را می دهند.

CDC چگونه کار می کند؟

هنگامی که CDC روی یک جدول فعال میشود، SQL Server از مکانیزم نامتقارنی (Async) استفاده می کند که به کمک آن تغییرات رخ داده در جداول را از فایل log می خواند و در جداولی که به منظور نگهداری تغییرات ایجاد کرده ذخیره میکند. در فایل های log هر رکوردی که ذخیره میشود یک شناسه یکتا با نام Log Sequence Number دارد که به اختصار LSN نامیده می شود. بسیاری از Metadata هایی که برای ما نگهداری می شود به خاطر قرابت نزدیکی که این تکنولوژی با Log دارد از همین اطلاعات Log استخراج می شوند.

برای فعال سازی CDC روی یک جدول ابتدا به کمک دستور `sys.sp_cdc_enable_db` باید آنرا در پایگاه داده مورد نظر فعال کنیم. با اجرای این دستور سیستم CDC بر روی پایگاه داده ما فعال می شود و زیرساخت های مورد نیاز آن مانند meta-data table ها، CDC Schema، CDC Database User و ... روی پایگاه داده ما ایجاد می شوند. در صورتی که نیاز داشته باشید می توانید در `sys.databases` و به کمک ستون `is_cdc_enabled` بررسی کنید که آیا CDC برای دیتابیس شما فعال است یا خیر.

حالا شما میتوانید از دستور `sys.sp_cdc_enable_table` برای فعال کردن CDC روی جدول خود استفاده کنید. با بررسی ستون `is_tracked_by_cdc` در `sys.tables` می توانید از فعال بودن یا نبودن CDC روی جدول خود مطلع شوید.

با فعال شدن این امکان، با هر تغییری در داده های جدول موردنظر شما، تغییرات آن در جدول CDC متناظر از Log استخراج شده و ذخیره می گردد. به صورت پیش فرض تمامی ستون های جدول نگهداری می شوند اما در صورت نیاز مثلا به دلایل امنیتی یا پرفورمنسی شما می توانید تنها اطلاعات بعضی از ستون های جدول را نگهداری کنید. این کار را به کمک پارامتر `captured_column_list@filegroup_name@` روی ایجاد CDC انجام می دهیم. همچنین به صورت پیش فرض این اطلاعات روی `FileGroup` اصلی ما نگهداری می شوند که باز هم در صورت نیاز می توانید با استفاده از پارامتر `filegroup_name@` یک `FileGroup` دیگر را جهت نگهداری اطلاعات معرفی کنید.

فعال بودن `SQL Server Agent` برای فعال کردن این امکان اجباری نیست. اما فعال بودن آن برای کارکرد صحیح CDC لازم است. به همین منظور در صورتی که هنگام فعال کردن این امکان روی جدول اگر این `SQL Server Agent` استارت نباشد، با اینکه کار فعال سازی درست انجام می شود اما به شما پیامی مبنی بر عدم کارکرد صحیح سیستم داده می شود.

آیا جدول CDC بسیار بزرگ می شود؟

با توجه به اینکه هر تغییراتی که در جداول ما اتفاق می افتد در جدول CDC نگهداری میشود، احتمالا باید این سوال برای شما پیش آمده باشد که آیا این جدول بسیار بزرگ نمیشود؟! پاسخ این است خیر. یک روال سیستمی وجود دارد که به صورت اتوماتیک هر ۳ روز یکبار داده ها قدیمی را پاک میکند. البته می توانید این اندازه را به هر اندازه ای که نیاز دارید تغییر دهید و یا می توانید به کمک روال ذخیره شده `sys.sp_cdc_cleanup_change_table` این کار را به صورت دستی انجام دهید. در صورت تغییر در جدولی که تغییرات آن را نگهداری می کنیم چه اتفاقی می افتد؟

با تغییر در ساختار جدول روال CDC به کار خود ادامه میدهد، با این تفاوت که در صورتی که ستونی اضافه شود، دیگر تغییرات آن نگهداری نمی شود، و اگر ستونی هم حذف شود مقدار `NULL` برای آن نگهداری می شود. یعنی CDC توجهی به تغییرات اتفاق افتاده ندارد و کار خود را به همان شکل سابق ادامه می دهد. در صورت نیاز شما می توانید یک روال دیگر برای نگهداری تغییرات ستون های جدید ایجاد کنید. اما به خاطر داشته باشید که تنها ۲ نمونه از نگهداری سابقه را می توانید برای یک جدول فعال کنیم.

خلاصه

در این مطلب، شرح کلی از CDC و نحوه فعال سازی آن را بر روی یک جدول داشتیم. در قسمت بعد با جزئیات این قابلیت جدید و نحوه استفاده از آن بیشتر آشنا می شویم.