



Miten Qlik Replicate mahdollistaa, että tietovaraston data pysyy reaaliaikaisena?

Change Data Capture (CDC) on Qlik Replicatteen sisäänrakennettu teknologia, jonka avulla Qlik Replicate tietää, milloin ETL-prosessi (Extract, Transform, Load) tulee aloittaa. CDC tutkii, milloin dataan on tehty muutoksia ja tunnistaa muutokset ennen ja jälkeen datan tarkastelun.

Lähdetietokannat ovat tyypillisesti tosiaikaisia tietokantoja, joissa datamuutokset tapahtuvat. Kun muutoksia on riittävä määrä ennakkoon määritettyjen kriteerien perusteella, käynnistyy Qlik Replicatteen ETL-prosessi automaattisesti ja datapäivitykset vietään tietovarastoon. Näin tietovaraston data pysyy lähes reaaliaikaisena, eivätkä päivitykset eivät ole tiettyyn

kellonaikaan sidottuna. Qlik Replicatteen avulla voidaan myös tehdä tarvittavat datamuutokset, kuten palstojen uudelleen nimeäminen, tietokentän datatyyppin muokkaaminen ja rivien erottelu.

Qlik Replicate on siis automatisoitu ETL-dataputki, joka mahdollistaa datan replikaatiotehtävien konfiguroinnin, ajastamisen, monitoroinnin ja hallinnoinnin. Qlik Replicate vähentää eri tietolähteiden välillä tapahtuvan datan kopioinnin aikaa ja kustannuksia, intuitiivisen graafisen käyttöliittymänsä ja "low-code/no-code" lähestymistapansa johdosta. [Tutustu Qlik Replicate -työkaluun tarkemmin Olikin verkkosivuilla.](#)

[CDC-prosessi on sisäänrakennettu myös Microsoftin Data Factory -työkaluun. Tutustu Data Factory -pipelinen pystyttämiseen ja CDC-prosessiin täältä.](#)

ETL-prosessin toteuttaminen CDC-toiminnallisuuden avulla

Change Data Capture (CDC) voidaan määrittää tutkimaan lähdejärjestelmää ja viemään automaattisesti tieto muutoksista tietovarastoon neljällä eri tavalla:

1. CDC-prosessia käytetään tietokantojen päivämäärä- ja kellonaikakenttien tarkasteluun. CDC siirtää päivitettyt tiedot tietovarastoon, kun lähdejärjestelmiin on tullut uutta tai muutettua dataa.
2. Eroavaisuuksien tarkastelu, jolloin CDC tarkastelee rivejä eli aikaisemman ja myöhemmän tietokannan tilojen eroja.
3. Tietokannan "triggereiden" käyttö CDC:llä sarakkeissa tehtyjen muutosten tarkasteluun.

Vaihtoehtoisesti tarkastelu voidaan kohdistaa pelkästään tärkeimpiin tietoihin ja niitä koskeviin muutostyyppeihin (tietojen lisäys, päivitys tai poistaminen).

4. CDC määritetään tutkimaan transaktionaalista tietokantaa, josta selviävät kaikki transaktionaalilokissa tapahtuneet muutokset.

Tyypillisesti näillä malleilla data päivitetään kerran yössä, ja useimmissa tapauksissa se riittääkin täyttämään datan ajantasaisuudelle asetetut vaatimukset. Modernien, automatisaatiota hyödyntävien teknologioiden ansiosta myös lähes reaaliaikaiset päivitykset ovat mahdollisia.