

Nieuwe database-spinoff CWI ziet het licht

15 juli 2021 11:55 | Rik Sanders

Topic Datamanagement



CWI-onderzoekers Hannes Mühleisen en Mark Raasveldt

Het Centrum Wiskunde & Informatica (CWI) kondigt een nieuwe spinoff aan: DuckDB Labs. De startup biedt een opensource-databasemanagementsysteem gericht op efficiënte data-analyse. Momenteel wordt DuckDB ongeveer 100.000 keer per week gedownload.

CWI-onderzoekers Hannes Mühleisen en Mark Raasveldt hebben de nieuwe CWI-spinoff DuckDB Labs opgericht. Het bedrijf helpt bij de verdere ontwikkeling van het systeem en levert adviesdiensten.

In 2019 bracht het duo in de Database Architectures-onderzoeksgroep van het CWI de eerste opensourceversie van [hun databasemanagementsysteem DuckDB](#) uit. Het is volgens de onderzoekers tevens het eerste speciaal gebouwde in-process Online Analytical Processing (OLAP)-database managementsysteem.

Drijfveer voor de ontwikkeling van zo'n systeem is dat bedrijven, overheden en academische onderzoeksgroepen steeds meer gegevens verzamelen. Die worden opgeslagen in grote databanken en een van de belangrijkste uitdagingen is om zo snel mogelijk nieuwe inzichten uit deze data te halen. Dat is een van de taken van een databasemanagementsysteem.

Eind

DuckDB, vernoemd naar een eend die Mühleisen als huisdier had, is twee jaar later een succes geworden: het wordt zo'n 100.000 keer per week gedownload, voornamelijk door datawetenschappers en zakelijke gebruikers. Door het in een apart bedrijf onder te brengen verachten Mühleisen en Raasveldt meer ruimte te krijgen om het systeem verder te ontwikkelen en ondersteunende [diensten](#) aan te bieden. Zij benadrukken dat DuckDB nog steeds een opensourceproject blijft onder de huidige 'MIT permissive free software license'.

DuckDB Labs is de nieuwste aanwinst in de lijst van nu 28 spinoffs van het CWI. Voorbeelden van eerdere afsplitsingen zijn Swatengineering, Datasplex, Stokhos, MonetDB Solutions en VectorWise.

Waarin onderscheidt DuckDB zich van bestaande database managementsystemen?

- DuckDB zich op analytische toepassingen waarin het nodig is om naar veel gegevens tegelijk te kijken. Denk aan gevallen waarin miljoenen rijen moeten worden geaggregeerd, of waarin gigantische tabellen moeten worden gecombineerd.
- DuckDB draait binnen andere processen die al op de computer draaien. Dat heeft het voordeel dat de gegevensoverdracht snel gaat. Volgens de makers is DuckDB het eerste in-process OLAP-databasesysteem dat grote hoeveelheden data beheert. Zij noemen het product ook wel 'SQLite voor analytics'. SQLite is 's werelds populairste database managementsysteem, maar het maakt geen analyses.
- DuckDB is een klein softwarepakket dat iedereen gemakkelijk kan installeren en waarvoor geen aparte server nodig is. Het systeem werkt snel omdat DuckDB voortbouwt op state-of-the-art database-onderzoek dat afkomstig is van de CWI Database Architectuur-groep. Het maakt gebruik van een verwerkingstechniek van zoekopdrachten, genaamd vectorized execution, die in 2005 werd ontwikkeld op het CWI.