

16-07-2021

DuckDB Labs nieuw CWI spin-off-bedrijf



Deel dit bericht



CWI-onderzoekers Hannes Mühleisen en Mark Raasveldt hebben het nieuwe CWI spin-off-bedrijf DuckDB Labs opgericht, dat DuckDB verder helpt ontwikkelen en consultancydiensten levert. DuckDB is een open source database managementsysteem gericht op efficiënte data-analyse. Het is gemakkelijk te installeren, werkt zeer snel en binnen lopende processen. Momenteel wordt DuckDB ongeveer 100.000 keer per week gedownload.

Bedrijven, overheden en academische onderzoeksgroepen verzamelen steeds meer gegevens. Deze gegevens worden opgeslagen in grote databanken en een van de belangrijkste uitdagingen is om zo snel mogelijk nieuwe inzichten uit de gegevens te halen. Dat is een van de taken van een database managementsysteem.

In 2019 brachten Hannes Mühleisen en Mark Raasveldt, onderzoekers in de Database Architectures onderzoeksgroep van het Centrum Wiskunde & Informatica, de eerste open source versie van hun database managementsysteem DuckDB uit. DuckDB is het eerste speciaal gebouwde in-process Online Analytical Processing (OLAP) database managementsysteem. "DuckDB dankt zijn naam aan het feit dat ik vroeger een eend als huisdier had", lacht Mühleisen. "Eenden zijn verbazingwekkende dieren. Ze kunnen vliegen, lopen en zwemmen, en ze zijn goed bestand tegen uitdagingen in de omgeving. Daarom is een eend de perfecte mascotte voor een veelzijdig en veerkrachtig data-managementsysteem."

Nu, twee jaar later, is DuckDB een groot succes geworden: het wordt zo'n 100.000 keer per week gedownload, voornamelijk door datawetenschappers en zakelijke gebruikers. "In een wereld waarin de succesvolste software vooral wordt ontwikkeld door Amerikaanse bedrijven, is het opmerkelijk dat software afkomstig van het door de overheid gefinancierde onderzoeksinstituut CWI zo'n vlucht heeft genomen", zegt Mühleisen.

Met als doel om een nog beter database managementsysteem te maken, hebben Mühleisen en Raasveldt recent DuckDB Labs B.V. opgericht als spin-off bedrijf van het CWI. Het bedrijf zal fungeren als omgeving voor innovatieve projecten rond DuckDB, om het systeem verder te ontwikkelen en als platform voor ondersteunende diensten. Mühleisen benadrukt dat DuckDB nog steeds een open source project blijft onder de huidige MIT permissive free software license.

Wat onderscheidt DuckDB van bestaande database managementsystemen? Raasveldt: “Allereerst richt 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. Er zijn veel van dit soort toepassingen in de zakelijke wereld en in statistische analyse.”

Ten tweede draait DuckDB binnen andere processen die al op de computer draaien. Raasveldt: “Als je data-analyse in Python doet, draait DuckDB binnen Python. Dat heeft het voordeel dat de gegevensoverdracht snel gaat. Eigenlijk is DuckDB het eerste in-process OLAP-databasesysteem dat grote hoeveelheden data beheert. Wij noemen onszelf de ‘SQLite voor analytics’. SQLite is ’s werelds populairste database managementsysteem, maar het maakt geen analyses.”

Vanuit praktisch oogpunt is DuckDB slank en doelgericht. Het is een klein softwarepakket dat iedereen gemakkelijk kan installeren en waarvoor geen aparte server nodig is. Tot slot werkt DuckDB snel. Dat komt 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.