

DATABASE- ONDERZOEK EN VOORLOPEN OP DE TOEKOMST

WAAR EEN KLEIN
LAND GROOT MAAR TE
BESCHIEDEN IN IS

Martin Kersten praat nuchter over het vooruitstrevende ict-werk van onderzoeksinstituut CWI, maar de Nederlandse wetenschapper heeft wel wéér een prestigieuze internationale prijs op zijn naam staan. 'Nee, we doen niet zoveel aan borstklopperij.'

TEKST: JASPER BAKKER BEELD: CWI

Eind juni vindt pas de officiële en feestelijke overhandiging plaats, maar de toekenning van de ACM Fellow 2016 award is eind vorig jaar al bekendgemaakt. Eén van de ontvangers van deze prestigieuze prijs is de Nederlandse wetenschapper Martin Kersten, verbonden aan het CWI (Centrum Wiskunde en Informatica). Een bijzondere prestatie, die slechts door enkele andere Nederlanders is gehaald. De bekendste hiervan zijn de vermaarde computerwetenschappers Edsger Dijkstra en Andrew Tanenbaum, die in respectievelijk 1994 en 1996 de Fellowship-award hebben gekregen. Het Nederlandse aandeel is dus een zeldzaamheid te noemen. Sterker nog: het internationaal gerenommeerde ACM (Association of Computing Machinery) blijkt qua award-toekenning nogal Amerikaans georiënteerd. Van de vijftig Fellowships

die voor afgelopen jaar zijn toegekend, zijn er drieënveertig voor Amerikaanse wetenschappers. De positie van Kersten in deze wetenschappelijke erkenninglijst is dan ook veelzeggend.

VLAGGENDRAGER

'Het is goed voor de bekendheid', erkent Kersten. Daarmee heeft hij het niet zozeer over hemzelf, maar over het CWI en Nederland. Kersten noemt zich nu wel 'vlaggendrager op het internationale podium'. Hij erkent dat ons land zichzelf best beter mag 'uitdragen' wat ict-onderzoek, -ontwikkeling en -vooruitgang betreft. 'Nee, we doen niet zoveel aan borstklopperij', omschrijft hij het gebrek aan vaderlandse marketing voor ict-innovatie. 'Amerikanen zijn daar beter in.' En wij mogen daar best een voorbeeld aan nemen. De bekende Nederlandse bescheidenheid is namelijk onterecht en kan zelfs een remmende werking

hebben. De toekenning van dit soort prijzen betekent namelijk erkenning voor verrichte inspanningen: onderzoekswerk en innovatie. En erkenning geeft weer naamsbekendheid, aan personen maar ook aan instituten en landen. Dat op zijn beurt trekt weer meer talent aan, wat volgende vooruitgangen aanjaagt. Een vliegwielwerking dus.

OOG VOOR DE MARKT

Het Nederlandse CWI, waar Kersten sinds 1985 aan is verbonden, heeft op dit gebied al een indrukwekkende staat van dienst. 'In de vijftiger jaren is hier de eerste computer gebouwd. En programmeertaal Python is hier gemaakt, door Guido van Rossum', somt hij enkele mijlpalen op. Belangrijk hierbij is dat het CWI meer is dan sec een onderzoeksinstituut. Het doet aan wetenschappelijk onderzoek met oog voor praktische toepassing. 'We kijken naar de lange termijn, mét overdracht naar de markt.'

Zo heeft het CWI in totaal 24 spin-off-bedrijven voortgebracht. Het gaat om onderzoek én ook transfer; van kennis, inzichten en technologie. Vandaar ook dat Kerstens werk niet beperkt is tot de wetenschappelijke wereld. Zijn onderzoek op databasegebied heeft al impact in de ict-wereld. MonetDB is daarvan het 'levende' bewijs: een opensourcedatabase gebaseerd op kolommen, in plaats van op rijen.

Kerstens ACM-erkenning is dan ook voor zijn ontwerp en implementatie van 'column store and main memory database systems'. Maar voordat hij de ACM Fellow 2016-award kreeg, had hij twee jaar terug al de ACM SIGMOD Edgar F. Codd innovations Award en vorig jaar de ACM SIGMOD Systems Award ontvangen. Beide andere prestigieuze prijzen waren ook voor kolom-gebaseerde databasetechnologie. Technische materie met praktische impact en concrete businessvoordelen. Dit andere database-ontwerp geeft namelijk een hoog prestatieniveau voor complexe queries op grote, grotere en enorme databases. Denk aan het combineren van tabellen die honderden kolommen en miljoenen rijen hebben. Vroeger waren dit haast onvoorstelbaar grote en ook wel zeldzame databases.

GROENTESOEP

Tegenwoordig is dat wel anders, mede dankzij dalende prijzen voor opslag, toenemende behoefte aan data, groeiende aantallen databronnen, en

Professor **Martin Kersten** is zijn ICT-onderzoekscarrière in 1975 begonnen als onderzoeksassistent. Slechts vier jaar later heeft hij het al geschopt tot onderzoeker en lector aan de Vrije Universiteit Amsterdam. Tot 1985 is zijn werk gericht op databassecurity en programmeertalen voor databases, waarbij hij ook een relationeel DBMS heeft ontwikkeld dat in gebruik is genomen voor een CASE-omgeving (computer-aided software engineering). Tussentijds is hij nog visiting researcher geweest aan de Universiteit van Californië. Het jaar 1985 markeert zijn overstap naar het CWI, waar hij de Database Research Group heeft opgezet.

niet te vergeten de trends van big data plus data-analytics. Reguliere ict-vooruitgang blijkt niet altijd toereikend om dit te behappen. Rekenkracht en opslagcapaciteit nemen weliswaar toe, tegen dalende kosten, maar rauw vermogen is niet heiligmakend. Soms is juist een andere aanpak nodig. Zoals dus kolom-gebaseerde databases.

Kersten geeft Computable een spoedcollege over databasetechnologie. 'Hoe zou je een database beschrijven aan je moeder?', begint hij zijn uitleg. 'Het is als groentesoep maken.' Het gaat om de ingrediënten, om de volgorde waarin je die klaar maakt en in de pan doet. En pas tegen het einde is de smaak bekend, aldus de CWI-onderzoeker. 'Het gaat om het totaalbeeld.'

Toegegeven, aan het eind van het kookproces valt er met wat kruiden nog wel iets te nuanceren aan de smaak, maar fundamenteel is de voorbereiding bepalend voor het eindresultaat. Zo is het ook met databases: het ontwerp bepaalt de uitkomst. 'Databases zijn sterk architectuurgericht.'

Van origine gebruiken relationele databases rijen. 'Dat werkt goed als je interesse hebt in één regel binnen een tabel, zoals bijvoorbeeld een bankrekeningnummer.' Dit is dus de kern van OLTP (On-line Transaction Processing) waarbij database simpele bewerkingen snel kunnen uitvoeren. Daar 'tegenover' staat OLAP (On-line Analytical Processing) waar veel meer data en veel meer bewerkingen bij zijn betrokken. Denk aan datamining, datawarehouses, big data en data-analytics, maar ook de huidige ict-trend van *deep learning*.

VOOROP LOPEN

Kolommen bewijzen hiervoor hun waarde. Kersten is daar al sinds begin jaren negentig serieus mee bezig, met 'voorwerk' aan gedistribueerde geheugensystemen en in-memory-databases dat zijn origine in de jaren tachtig heeft. Jawel, datamining en in-memory-databases waren bij het CWI al aan de orde voordat automatisering en databases doordrongen in de algemene zakenwereld.

'De trend was helder', zegt Kersten over dat vroege onderzoekswerk. Daarbij zijn toen wel praktische problemen opgedoken, waaronder performance die onder de maat was voor real-world-gebruik. Sinds eind jaren negentig is een veelbelovend experimenteel platform verder verfijnd. Begin deze eeuw is daaruit de opensourcedatabase MonetDB voortgekomen. 'In 2005 was het redelijk volwas-





WE KIJKEN NAAR DE LANGE TERMIJN, MÉT OVERDRACHT NAAR DE MARKT

te ontsluiten voor gewone astronomen. Zij hoefden niet langer computerwetenschapper te zijn om de data te benutten.

‘Wij gingen dus die bestaande database-applicatie implementeren in ons systeem: MonetDB.’ Een flinke opgave waarbij nog ontwikkelwerk aan de eigen database nodig bleek. ‘We hebben er ongeveer tweeënhalf jaar aan gewerkt.’ Het eerste halfjaar was voor aanpassingen aan MonetDB zelf. Uiteindelijk heeft al dit werk vruchten afgeworpen, ook voor databasesoftware zelf, wat dankzij de opensource-aard weer toegankelijk is voor iedereen.

HOOGTEKAART, DAKEN EN ZONNEPANELEN

De toepassingen gaan dan ook verder dan alleen het wetenschappelijke veld van astronomie.

Praktische toepassingen zitten ook in vergezocht lijkende hoeken. Kersten vertelt dat er afgelopen drie jaar 17 TB aan LIDAR-data (Laser Imaging Detection And Ranging) is vergaard waarbij lasers vanuit vliegtuigen accurate hoogtebepalingen hebben gedaan. Met wel tien meetpunten per vierkante meter komt de resolutie van deze hoogtekaart uit op enkele centimeters. ‘Dus een auto is zichtbaar, maar een bal in de tuin ook.’

Met dergelijk gedetailleerde data is veel te doen, wetenschappelijk maar ook heel alledaags. Een voorbeeld dicht bij huis is de bepaling van welke daken geschikt - en welke het meest geschikt - zijn voor de plaatsing van zonnepanelen. Deze verrassend praktische toepassing is net als vele andere mogelijk dankzij jaren aan wetenschappelijk werk aan het CWI. MonetDB en Kerstens ACM Fellowship zijn daaruit voortgekomen wapenfeiten, naast relatief onbekende prestaties als Data Distilleries, VectorWise en meer. ■

sen, dachten we’, vertelt Kersten.

De relaterende bijzin komt voort uit de praktijkervaring van een ambitieus project om de waarde van MonetDB aan te tonen. Dit was het implementeren in deze nieuwkomer van de omvangrijke sql-database van astronomiecatalogus Sloan Digital Sky Survey (SDSS). Die database was gemaakt door de gerenommeerde computerwetenschappers Jim Gray en Alexander Szalay. ‘Zij hebben een pivot in de astronomie veroorzaakt’, door de enorme datahoeveelheden van het SDSS

