



door Jos Baeten

Jos Baeten is algemeen directeur Centrum Wiskunde & Informatica (CWI), sinds 1946 het nationale onderzoeksinstituut voor wiskunde en informatica. Het is deel van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). Het instituut heeft een sterke internationale positie. Ruim 150 wetenschappers doen er grensverleggend onderzoek en dragen de verkregen kennis over aan de maatschappij. Meer dan 30 van de onderzoekers zijn hoogleraar aan een universiteit. Het instituut heeft 22 spin-off bedrijven voortgebracht.
www.cwi.nl

Big data, de hype voorbij

Big data zijn overal; big data zijn een hype. De ICT sector is vol van hypes. Was het eerst SaaS wat de klok sloeg, toen werd het agile en scrum waar iedereen

achteraan holde. En nu dus big data. Ook de overheid en de onderzoeksfinanciers doen er lustig aan mee. Alles heet big data - ook als het niet gaat om vele exabytes aan data die te veel, te snel en te divers zijn om met gewone databasemanagementsystemen te onderhouden of te doorzoeken. Ik durf wel te stellen dat de hype op dit moment groter is dan de behoefte.

Maar in ons onderzoek doen we niet aan hypes. Wij doen aan lange termijn. Al ruim dertig jaar houden we ons op het Centrum Wiskunde & Informatica (CWI) in Amsterdam bezig met big data onderzoek, toen het nog lang niet zo heette. En dus niet omdat het een hype is.

Ons databaseteam ontwikkelde vanaf de jaren tachtig visionaire en innovatieve manieren om reusachtige hoeveelheden informatie te ontsluiten, geschikt voor de verre toekomst - als gewone systemen al lang niet meer schaalbaar zijn. Het zette de databasemanagementwereld op zijn kop. Bijvoorbeeld door de informatie

niet in rijen maar in kolommen op te slaan. Door vaak gezochte informatie te clusteren. Door niet één vast format voor data-input te eisen maar flexibel te zijn. Door intelligent spelend nieuwe wegen uit te proberen. En door niet te kiezen voor snelle symptoombestrijding maar gefocust en doelgericht de oorzaken van de problemen te onderzoeken. Door diep te graven en te begrijpen. Door heel doordacht, in een open omgeving, onze systemen verder te ontwikkelen. Gedreven door nieuwsgierigheid en niet uit effectbejag.

Het resultaat is dat ons supersnelle MonetDB databasemanagementsysteem nu overal ter wereld wordt gebruikt in toepassingen die de maatschappij en de economie verder helpen. Het wordt bijvoorbeeld ingezet bij onderzoek naar computercriminaliteit. Op in beslag genomen computers van criminelen staan vele files, waarvan de informatie met een standaard systeem niet te achterhalen is. De kunst is het om daar automatisch juist die data uit te halen die iets vertellen over drugs of kindermisbruik. Ons MonetDB systeem stond aan de basis van het XIRAF-systeem van het NFI, dat dit precies doet. MonetDB helpt ook bij het verwerken van data uit de gigantische Sloan-sterrendatabase. Maar ook bij de analyse van ziektekostenclaims en andere bedrijfstrends.

Ons werk blijft dus niet alleen bij fundamenteel onderzoek. Uit de databasegroep op het CWI zijn vier succesvolle spin-off bedrijven ontstaan. Een is er gespecialiseerd in consultancy voor business analytics, een ander is overgenomen door het Amerikaanse bedrijf Actian (voorheen Ingres), dat onze technieken in haar producten heeft overgenomen, net als overigens veel andere bedrijven.

Naast databasemanagementsystemen omvatten big data meer; input, visualisatie en business intelligence spelen ook een grote rol. Op dat laatste gebied onderzoeken wij op het CWI manieren om de aanrijtijden van ambulances in Nederland te verkorten. Met data-analyse en slimme wiskunde ontwikkelen we methodes om ongelukken als het ware te voorspellen en ambulances dynamisch zo neer te zetten dat ze het snelst bij de plaats zijn waar je - met

data-analyse uit het verleden - volgende incidenten verwacht. Ambulances kunnen zo veel efficiënter rijden.

De hele wereld - en vooral Nederland - willen we laten delen in deze resultaten en daarom stellen we veel van ons werk open source beschikbaar. Universiteiten maken onze succesvolle onderzoekers tot deeltijdhoogleraar, zodat via onderwijs nog meer jonge talenten van dit onderzoek kunnen profiteren. En ook Europa ziet het belang, getuige de vele Europese projecten. Verscheidene belangrijke internationale prijzen werden aan onze onderzoekers toegekend voor hun innovatieve en toonaangevende werk.

Door onze resultaten in big data vindt iedereen ons nu begeerlijk - maar beseft dat er dertig jaar investering achter zit. We kunnen niet voorspellen wat er over dertig jaar een hype is; wij werken op andere tijdschalen. Niemand kon voorspellen dat de programmeertaal Python, ontworpen in de jaren negentig op het CWI, jaren later wereldwijd miljoenen gebruikers zou hebben - van Google en NASA tot grote banken. Een Nederlands onderzoeksresultaat waar we best trots op kunnen zijn.

Hoewel big data nu vooral een hype zijn, is het doorzoeken en beheren van enorme hoeveelheden snel groeiende data in de toekomst wel een groot probleem. Een probleem waar Nederland in potentie de oplossing voor heeft, door onderzoekers vrij te laten. Aan iedereen wil ik zeggen: laat je niet meesleuren door de hype maar kijk wat erachter zit. Zie het belang van lange-termijnonderzoek en investeer erin. Dan blijven big data geen hype maar een realistisch middel om de maatschappij blijvend verder te helpen.

Big data dragen bij aan het vertrouwen van de burger

“De komende decennia zal de wereldeconomie razendsnel en ingrijpend veranderen. Het gaat om economische machtsverschuivingen in de wereld tussen westerse industrielanden en opkomende economieën. Deze worden vooral opgeschud door de snelle opmars van digitalisering”, stelt Anwar Osseyran.

Osseyran schetst een beeld van start-ups die met een tablet, smartphone en website de beschikbare data in de clouds goed benutten en succesvol de concurrentieslag aangaan met bestaande, gevestigde bedrijven. Daardoor zijn de afgelopen jaren wereldwijd talloze traditionele bedrijven ten onder gegaan omdat ze zich niet snel genoeg hebben aangepast aan de markt. Denk aan de muziekindustrie, de reissector en de detailhandel. “Technologische ontwikkelingen zullen zowel een sterke invloed hebben op economieën als op de bedrijfsresultaten van een toenemend aantal bedrijven. Van cruciaal belang daarbij is dat bedrijven gebruik gaan maken van alle interne- en externe digitale informatie die relevant en beschikbaar is.”

“Internationaal onderzoek wijst uit dat veel ondernemers met een adequate analyse en de juiste toepassingen van deze big data hun concurrentiepositie kunnen versterken en zo hogere winst realiseren. Toch is het niet bepaald eenvoudig uit de wirwar van big data trends te filteren en daar waardevolle conclusies aan te verbinden. Wie daarin slaagt, heeft de sleutel tot onze op data-gebaseerde toekomst”, aldus Osseyran.

Vertrouwen van de burger

Hij stelt dat in de openbare sector een schat aan gegevens wordt verzameld. “Bijvoorbeeld op het gebied van energie, gezondheid, belastingen en sociale zekerheid, sociodemografische informatie en opiniepeilingen. Eén van de aspecten waarvoor in de overheidssector dringend verbetering nodig is en waar big data uitkomst kunnen bieden, is transparantie en het vertrouwen van de burger.”



Anwar Osseyran is algemeen directeur van SURFsara (het nationale High Performance Computing Center in Nederland), Hoogleraar Business Analytics and Computer Science aan de Universiteit van Amsterdam (UVA) en bestuurslid van diverse ICT-bedrijven en organisaties.

“De beschikbaarstelling van de enorme hoeveelheid informatie als open data zal de transparantie ten goede komen en de burger meer duidelijkheid geven over welke gegevens de overheid verzamelt en wat zij daarmee doet. Meer transparantie betekent dat de burger de overheidsbestedingen kan volgen en overheden

“De beschikbaarstelling van de enorme hoeveelheid informatie als open data zal de transparantie ten goede komen.”

dwingt tot doelmatigheidsverbetering en kostenverlaging. Dat zou overheden stimuleren de burgers te respecteren en publiek-private initiatieven aanmoedigen innovatieve diensten te ontwikkelen die beter voorzien in de behoeften van de samenleving.”

Zoals op veel andere technologische gebieden zal de daadwerkelijke impact van big data-analyse pas na verloop van tijd duidelijk worden, zegt Anwar Osseyran. “De beloften van big data zijn innovatie, groei en blijvende duurzaamheid. Risico's zijn inbreuk op de persoonlijke levenssfeer, eigendomsrechten, gegevensintegriteit en persoonlijke vrijheid, maar ook veiligheid. Bij een open en transparante inzet van big data is het dus onontkoombaar dat ook de keerzijde van big data aan het licht komt. Dus ook hier geldt: bezint eer ge begint...”