



Niels met twee oogstmachines die horizontale schaalbaarheid illustreren

KLANTEN INSPIREREN MET DATA

EEN BERG AAN CLICKS

Niels Basjes is IT-architect bij bol.com. Bol.com bevindt zich in de voorhoede van de gebruikers van big data. Hoe meer gegevens beschikbaar zijn, des te beter kan de klant online bediend worden.

Niels is één van de drie Lead IT-architecten, binnen een team van vijftien architecten die big data-oplossingen beschikbaar maken voor de business units. Bol.com bevindt zich in de voorhoede van de gebruikers van big data. 'Uiteraard zijn we wat volume betreft kleiner dan Google en Twitter, maar als het gaat om innovatie dan doen we zeker mee met de grote jongens. Wij zijn meer dan een eindgebruiker. Voor ons is big data geen bijproduct, maar een kernactiviteit. In Nederland zijn er misschien maar vijf partijen die zo veel uit big data halen als bol.com.' Bol.com kan razendsnel reageren met bijvoorbeeld instant productaanbevelingen. Door 'real-time complex event processing software' toe te passen, kan directe feedback aan de klant worden gegeven op basis van zijn gedrag op de site. Met zo'n 5 miljoen klanten en een product-

catalogus van vele miljoenen producten is het noodzakelijk om het personaliseren te automatiseren. Bij bol.com helpen deze technologieën om daadwerkelijk de persoonlijke winkel te realiseren. Er zijn echter zoveel data beschikbaar, dat er slimme systemen nodig zijn om de data gebruiksklaar te maken voor commerciële doeleinden.

KENNIS IS SCHAARS

De kennis en ervaring over de productie en het effectief inzetten van gegevens zijn nog schaars. Daarom traint Niels systeemontwikkelaars binnen en buiten het bedrijf. Hij geeft lezingen en gastcolleges en weet als geen ander de vertaalslag te maken van techniek naar business. 'Ik verzin nieuwe IT-systemen', vertelt

'DE BELANGRIJKSTE VRAAG IS: HEEFT DEZE ZOEKTERM DE KLANT GEHOLPEN?'

hij. Hij geeft er de voorkeur aan om het oorspronkelijke probleem 'op te blazen' en het dan weer terug te brengen tot iets waarvoor de oplossing technisch effectiever is. 'Wanneer bijvoorbeeld een businessafdeling een wens heeft, wordt die wens vaak klein geformuleerd omdat men bang is dat er geen oplossing is voor wat men werkelijk wil. Mensen proberen hun vraag in te kaderen vanuit de gedachte dat de oplossing dan makkelijker te vinden is. Maar ik ga juist vaak nog wat verder dan wat de collega dacht te willen.' Hij probeert vervolgens om er een IT-oplossing voor te ontwikkelen. 'Ik was na de TU Delft een echte computerwetenschapper', vertelt hij. 'Bij het Nationaal Lucht- en Ruimtevaartlaboratorium (NLR) begreep ik dat ik op een heel smalle, erg technische basis opereerde. Ik begreep mensen die geen infra-architect waren vaak niet en zij mij niet. Er ontbrak iets.' Hij volgde Nyenrode PDP4, en die opleiding bleek het ontbrekende puzzelstukje. Zijn medestudenten, die geen technische achtergrond hadden, gaven hem het inzicht in de grenzen van het kennisniveau. Niels schakelt snel, zeer snel, en moest geduld ontwikkelen om zijn visie helder te maken. Met een aanvullende NLP-opleiding ontwikkelde hij de vaardigheden voor effectieve communicatie. Zo leerde hij nieuwe concepten uit te leggen en dóór te vragen: wat wil je bereiken met je campagne? Wat is je doel? Daardoor is zijn rol nu meer hybride: op het raakvlak van IT en business.

OPEN SOURCE

Niels en zijn collega's borduren vaak voort op door Google en LinkedIn ontwikkelde open source-software. De internetgiganten

maken deze software doelbewust open source, zodat gebruikers zoals Niels de software verder kunnen verbeteren. Daardoor heeft hij al aan innovaties bij LinkedIn en Twitter kunnen bijdragen. In eigen huis gaan de systemen pas live na een uitgebreide testperiode. 'We hebben een lab opgezet waarmee we met de software kunnen experimenteren. Zo kunnen we een server stil zetten en de piekbelasting van de programmatuur testen. Pas na gebleken geschiktheid gaan we live', zegt Niels.

Hij gebruikt onder andere Hadoop. Hadoop helpt bij horizontale schaalbaarheid. 'Big data zijn net immense velden met sojabonen. Die velden kun je oogsten met één megamachine, maar dat is duur en inefficiënt. Zo'n machine, hoe groot ook, doet er veel langer over dan een reeks veel kleinere en goedkopere oogstmachines die in een V-vorm over de velden gaan. Dat betekent dat de prestaties en de capaciteit min of meer lineair toenemen met het aantal servers dat wordt gebruikt. De gedachte is dan ook om het probleem in kleine onafhankelijke brokken te hakken en te verspreiden over veel computers, om vervolgens de oplossingen te integreren', legt Niels uit. Daardoor kan commodity hardware worden gebruikt: kleine, goedkope systemen die makkelijk en snel vervangen kunnen worden. 'De kosten voor het verwerken van de gegevens lopen enorm op zonder die schaalbaarheid. Het is zoals de curve van een hockeystick, een twee keer zo krachtige server kost vaak meer dan twee keer zo veel. Op deze manier kunnen we heel eenvoudig door blijven groeien in de beschikbare rekenkracht. Wij schalen bij met extra processoren zonder dat de kosten exponentieel toenemen en zonder de software te hoeven aanpassen. We winnen enorm veel tijd. Daardoor kunnen we de klant nu bieden wat tien jaar geleden voor onmogelijk werd gehouden. Neem de zoekterm Harry Potter. Al bij het typen van 'har' ziet de klant suggesties voor categorieën zoals hardlopen in Sport & Vrije tijd en een hele reeks zoeksuggesties van, inderdaad, Harry Potter tot DJ Hardwell. We leveren meteen de populaire zoekresultaten met bijvoorbeeld een Harry Potter DVD reeks. Na slechts drie toetsaanslagen', vertelt Niels. Maar daar houdt het niet op: 'De belangrijkste vraag is: heeft deze zoekterm de klant geholpen? De meest succesvolle zoektermen zetten we boven aan in het lijstje. De mindere zoektermen worden handmatig geanalyseerd: wat moet er gebeuren om de volgende keer ook deze klant te laten slagen?' Uiteraard wil bol.com klanten inspireren en is het bedrijf in staat juist die producten te vinden waar hij of zijn niet direct aan gedacht had. Niels: 'Een leuk voorbeeld vind ik dat bij sommige soorten technisch spelgoed ook relevante boeken als aanbevelingen worden getoond. Dat zijn patronen die als vanzelf zijn komen boven drijven en dus met onze big data-oplossingen zijn ontdekt in de gigantische berg van vele miljarden clicks.'

WIE IS...

NIELS BASJES
(20000302)

WERK

(Lead) IT Architect, bol.com vanaf 2008
Sr Web Analytics Architect, Moniforce 2005-2008
ICT Architect, NLR 1997-2004
Software Engineer, USoft 1995-1997

OPLEIDING(EN)

Nyenrode PDP 2000-2003
TU Delft MSc Technical Computer Science 1989-1995

nbasjes@bol.com