Publicatie : Quest bijlage
Datum : 31 dec 2013
cm2 : 2.827
Pagina : 52
Advertentiewaarde : onbekend
Regio : Nederland
Frequentie : incidenteel
Oplage : 183.348

BIG DATA

Supercomputers, servers en rekentrucs scheppen orde in enorme bergen gegevens

Veel, meer, meest

Spitten in hopen onderzoeksgegevens, zoeken naar naalden in hooibergen. Wetenschappers gebruiken steeds vaker enorme databases en krachtige computers bij hun onderzoek. NWO helpt hen met de noodzakelijke infrastructuur.

TEKST: DAVID REDEKER

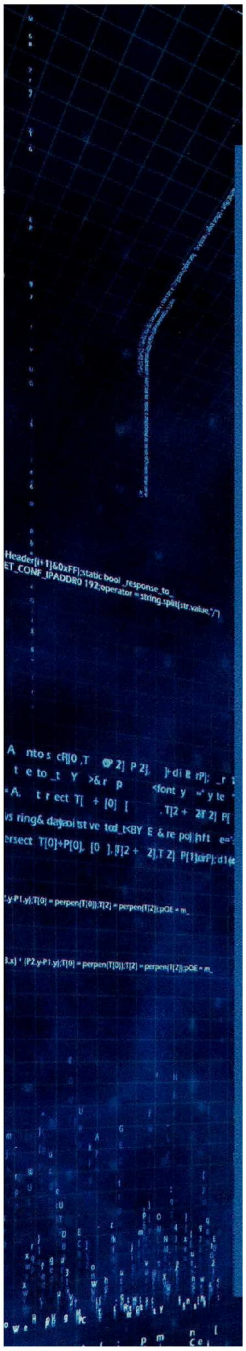
Alle data verzamelen

Een centrale supercomputer en een netwerk van honderden snelle processors die samen berekeningen uitvoeren. Voor een individuele onderzoeker of onderzoeksgroep zijn deze faciliteiten onbetaalbaar. Daarom heeft Nederland een aantal apparaten en diensten verzameld voor de wetenschap. Op deze pagina's staan 6 voorbeelden van onderzoeken die daar gebruik van maken. Het hart wordt door het BiG Grid-project gevormd, dat van 2006 tot 2012 liep en gehost werd door het Nikhef (het Nationaal instituut voor subatomaire fysica) en SARA . Door dat project is er nu een zogeheten 'e-infrastructuur' van computernetwerken en -systemen. Want hoewel het project voorbij is, leven de faciliteiten bij SURFsara voort.

Bloedstollend mooi model

Wat is het? Een computerprogramma dat na kan bootsen hoe bloed stolt en hoe het door aderen stroomt.
Waarom bijzonder? Dit model brengt alle details samen die in de afgelopen decennia zijn ontdekt over het stollen en stromen van bloed. De oude modellen stonden te ver af van de werkelijkheid. Nu kunnen onderzoekers naar hartelust virtueel aan de knoppen van de bloedstolling draaien.
Onderzoek? Als je in je vinger snijdt, bloed je niet dood. Je lichaam heeft een ingenieus systeem dat ervoor zorgt dat het bloed stroomt waar het kan, maar stolt waar het moet. Biologen proberen al decennia de vinger achter dat systeem te krijgen. Maar het was te ingewikkeld, want tientallen verschillende eiwitten grijpen op verschillende momenten op elkaar in. Een vormverandering van eiwit nummer één zet de productie van nummer 34 in gang die dan weer nummer 12 blokkeert en nummer 64 en 23 een duwtje geeft. Het was om moedeloos van te worden. Bart Bakker en Henk van Ooijen van de Life Science

Publicatie : Quest bijlage
Datum : 31 dec 2013
cm2 : 2.827
Pagina : 52
Advertentiewaarde : onbekend
Regio : Nederland
Frequentie : incidenteel
Oplage : 183.348



Hoe Wikipedia het wiel uitvond

Wat is het? Onderzoek naar hoe het trefwoordsysteem van Wikipedia evolueert.
Waarom bijzonder? Na economen, biologen en natuurkundigen ontdekken nu ook geesteswetenschappers de kracht van computers, enorme datastromen en gespreid berekenen.
Onderzoek? 100 jaar geleden bedacht de Belg Paul Otlet een systeem dat alle bibliotheekboeken en tijdschriftartikelen indeelt op een overzichtelijke manier. Toen Wikipedia in 2001 werd opgericht dachten encyclopediemakers, bibliothecarissen en archivarissen dan ook dat Wikipedia het Belgische systeem wel zou kopiëren. Maar nee, deze online encyclopedie wilde werkenderwijs vanzelf een rubricering opbouwen. Dat is gelukt, zegt Andrea Scharnhorst van DANS (Data Archiving and Networked Services), een gezamenlijk instituut van NWO en KNAW dat digitale archieven toegankelijk maakt voor andere onderzoekers. Scharnhorst downloadde met haar team de complete Engelse Wikipedia van 2008 naar de servers van BiG Grid. En daarna gooide ze alle tekst van de 4 miljoen artikelen in de prullenbak. Ze bewaarde alleen de links tussen Wikipedia-artikelen en de links naar de trefwoordpagina's die onderaan de pagina staan. 'De trefwoorden van Wikipedia evolueren geleidelijk naar een systeem dat verbazingwekkend op de nu 100 jaar oude Universele Decimale Classificatie lijkt.'
Cijfer? 5 dagen. Zolang duurt het als je de Engelse Wikipedia downloadt via een internetverbinding thuis.
Nut? 'Onze methode kun je ook gebruiken om aan buitenstaanders in een oogopslag duidelijk te maken wat ze wel en wat ze niet kunnen vinden in een archief of naslagwerk', zegt Scharnhorst. 'En je kunt ook netwerken van personen en hun kennis in kaart brengen. De infographic die wij van ons onderzoek maakten, is de hele wereld overgegaan. Hij is tentoongesteld en we hebben er vervolgfinanciering door gekregen.'
Toekomst? 'Er komen steeds meer data', zegt Scharnhorst. 'Die moet je op de een of andere manier ordenen of rubriceren. Daarom hebben we een Europees project gestart waarin informatiewetenschappers met natuurkundigen samenwerken en visualiseren hoe kennis in elkaar zit.' Het project heet KnowEscape. Want vluchten voor kennis kan niet meer.

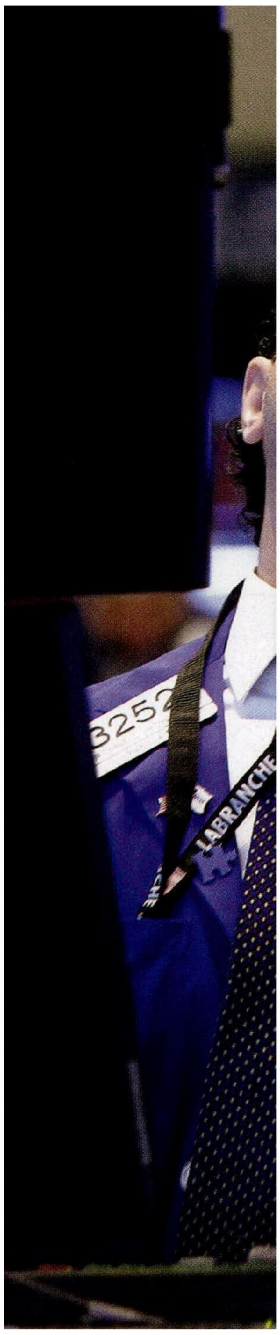
Het oude systeem van de bibliotheek lijkt sprekend op dat van Wikipedia.

Het lijkt zo simpel, een wondje dat heelt. Maar bloedstolling is complex.

Facilities van Philips Research verzamelden alle bestaande kennis in een overkoepelend computermodel. Het model was zo groot en ingewikkeld dat hun laptop er in een mum van tijd op vastliep. Gelukkig was daar BiG Grid.
Cijfer? Eén eeuw. 'Als we onze berekeningen op één computer zouden moeten uitvoeren, waren we 100 jaar bezig', zegt Van Ooijen. 'Met BiG Grid kon dit binnen een paar dagen.'
Nut? Bakker: 'We werken nu samen met klinisch epidemioloog Frits Rosendaal (die in 2012 een NWO-Spinozapremie kreeg, red.). Hij heeft een enorme database met patiënten die lijden aan trombose. Ons model geeft hem een nieuwe kijk op zijn gegevens.'
Toekomst? Van Ooijen: 'Philips zorgt in ziekenhuizen voor apparaten die patiënten monitoren. Met ons model kunnen we bijvoorbeeld bijdragen aan het verbeteren van de bewaking rondom operaties. En we willen specialisten helpen met hun keuze om de ene patiënt wel en de andere geen bloedverdunners te geven.'

Publicatie : Quest bijlage
Datum : 31 dec 2013
cm2 : 2.827
Pagina : 52
Advertentiewaarde : onbekend
Regio : Nederland
Frequentie : incidenteel
Oplage : 183.348

De brug van de toekomst geeft zelf een seintje als er reparaties nodig zijn

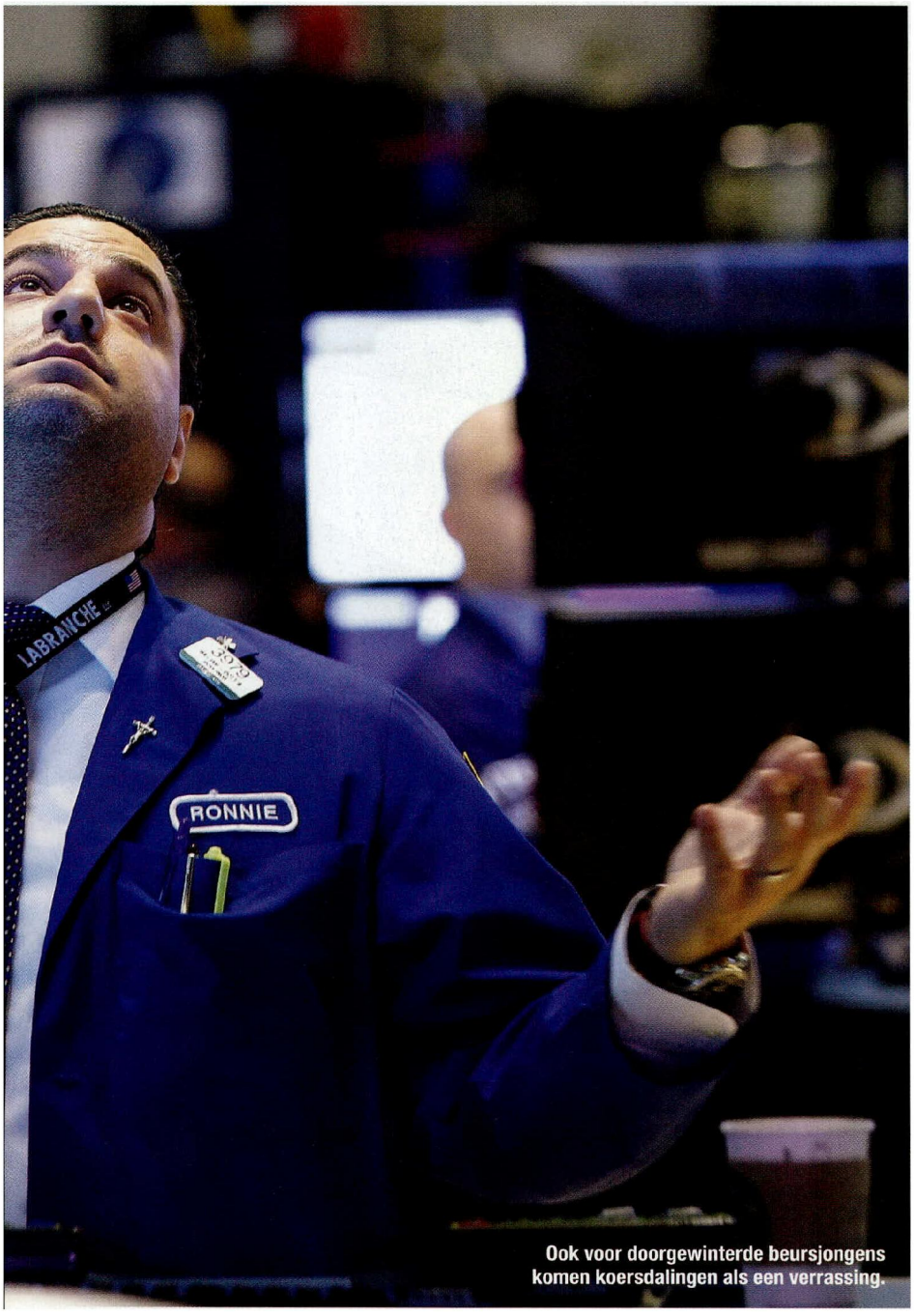


Brug vraagt hulp als het moet

Wat is het? Een intelligente brug die aangeeft hoe snel hij slijt en waar onderhoud nodig is.
Waarom bijzonder? Op de A6 tussen Amsterdam en Almere ligt een van de best bestudeerde bruggen ter wereld: 145 sensoren sturen continu gegevens over trillingen, temperatuur en trekkrachten naar de Universiteit Leiden.
Onderzoek? Bouwbedrijf Strukton monteerde tijdens de renovatie in 2007 en 2008 145 sensoren op de Hollandse Brug. Deze sensoren volgden de uitharding van het beton. Na de renovatie bleven ze zitten voor de wetenschap. Leidse onderzoekers willen ontdekken of de brug slijt en waar onderhoud nodig is. Het grote probleem is dat de brug nooit eens even lekker stil hangt. Er is nooit sprake van een door wetenschappers zo gewilde nulmeting. Een warme brug buigt bijvoorbeeld dieper door dan een koude brug. En een brug met zijwind wappert langer na, dan een brug bij windstil weer. Dus zelfs al zou één vrachtwagen over de brug rijden met voor en achter hem 10 minuten geen verkeer, dan nog gooien de temperatuur en de wind roet in de nulmeting. Maar speciale software en de vereende krachten van BiG Grid schiepen orde in de chaos.
Cijfer? 2 minuten. Dat is de tijd die het nu kost om 3 maanden aan bruggegevens op te schonen. Een paar jaar terug wisten de onderzoekers niet hoe ze alle data verwerken konden. Nu hebben ze software geschreven die 528 computers tegelijk aan de gegevens laat rekenen.
Nut? Joaquin Vanschoren van de Universiteit Leiden: 'We leren door de Hollandse Brug welke sensoren nuttig zijn, waar ze het beste hangen en hoe vaak ze moeten meten.' Er zijn, naast temperatuur- en trillingsmeters, bijvoorbeeld ook sensoren die in gewapend beton de geleiding meten. Zo zie je of een brug roest.
Toekomst? Honderden bruggen wachten met smart op de slimme sensoren. Die kunnen de plaats innemen van dure en ingrijpende inspecties waarbij de weg dicht moet. Daarna zijn ook gebouwen aan de beurt. En tunnels en lantarenpalen.

Onder het brugdek is in elk geval ruimte genoeg voor sensoren.

Publicatie : Quest bijlage
Datum : 31 dec 2013
cm2 : 2.827
Pagina : 52
Advertentiewaarde : onbekend
Regio : Nederland
Frequentie : incidenteel
Oplage : 183.348



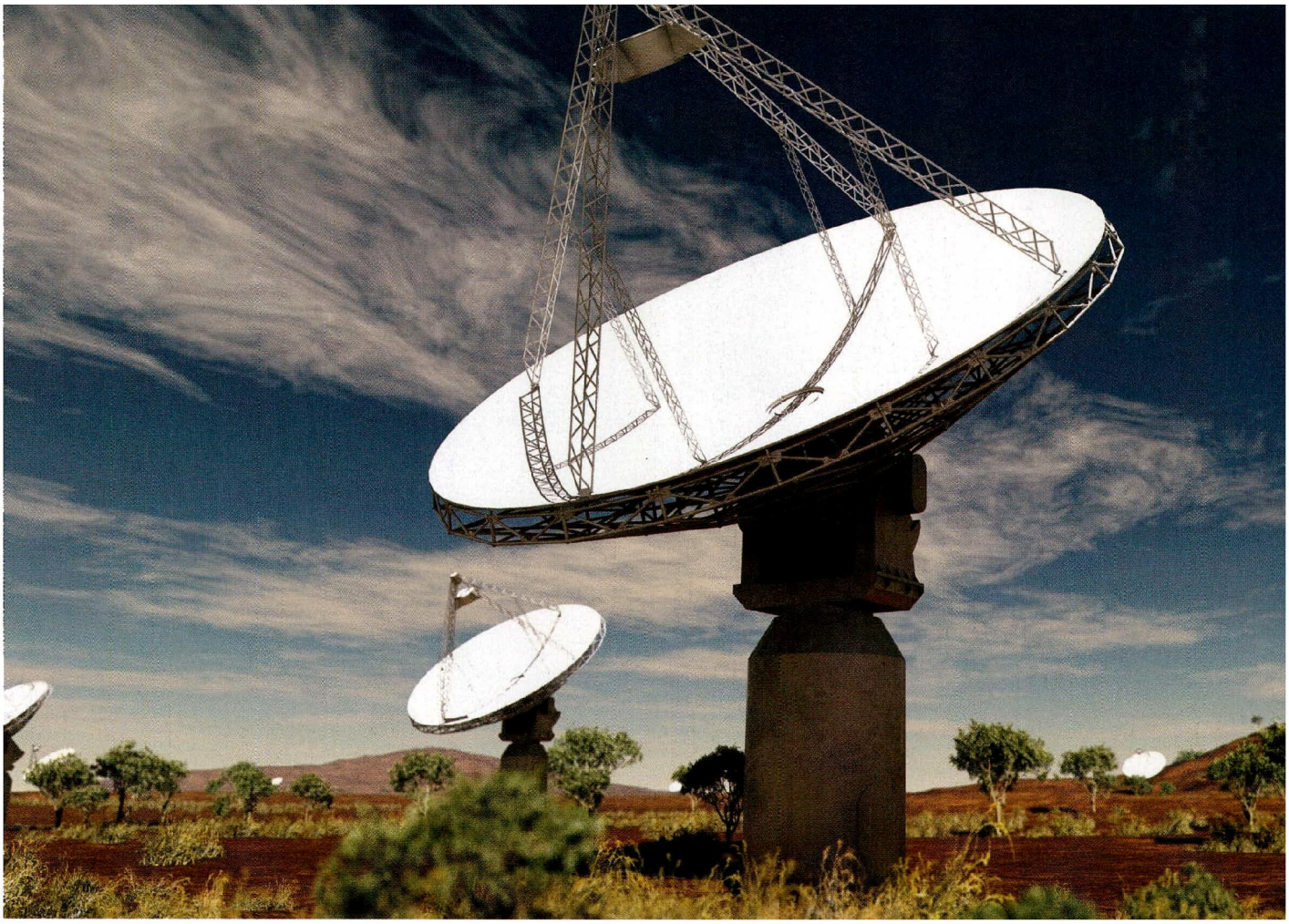
Ook voor doorgewinterde beursjongens komen koersdalingen als een verrassing.

Besmettelijke beurzen

Wat is het? Economen onderzoeken hoe de financiële crisis zich als een besmettelijke ziekte over de wereld verspreidde.
Waarom bijzonder? De wetenschappers hebben de beschikking over een gigantische database. Van 600 financiële markten is tot op de microseconde bekend hoeveel elk aandeel waard was. En dat gedurende 17 jaar.
Onderzoek? Mensen kun je inenten tegen ziekten. Een kippenstal met vogelgriep kun je isoleren van de buitenwereld. Maar als het misgaat op de beurs, verspreidt de malaise zich als een lopend vuurtje naar andere markten en andere landen. Hoe werkt dit? Waarom vallen sommige beurzen als dominostenen om en houden andere markten dapper stand? Dat zijn de vragen waar Mathijs van Dijk en zijn team aan de Rotterdam School of Management van de Erasmus Universiteit zich mee bezig houden. Van Dijk: 'We hebben 2 zeer grote databases die we naast elkaar leggen op BiG Grid. De ene database bevat transacties, de andere prijzen. Als je die 2 combineert dan kun je per transactie afleiden of het een verkooptransactie is of een kooptransactie. En dat geeft weer informatie over het vertrouwen van handelaren in de beurs.' De onderzoekers nemen nu 40 van de 600 markten in de database onder de loep. Het liefst zouden ze de 600 markten allemaal bekijken, maar dat is niet haalbaar vanwege de enorme hoeveelheid van zo'n 2 petabytes aan gegevens.
Cijfer? 0,3 seconde. Zolang duurt het om met je ogen te knipperen. In die tijd kopen en verkopen automatische handelssystemen miljoenen aandelen. 'Er wordt vaak met een beschuldigende vinger gewezen naar die *high frequency traders*', zegt Van Dijk. 'Maar voorlopig hebben we geen bewijs dat zij verantwoordelijk zijn voor het omvallen van de beurzen.'
Nut? Als je weet waarom sommige beurzen wel standhouden tijdens een crisis, kun je andere beurzen ook weerbaarder maken. Bijvoorbeeld met regels over het tijdelijk stilleggen van de handel. Ook grote beleggers, zoals de pensioenfondsen, kunnen hun strategie aanpassen zodat ze minder risico lopen. Alles bij elkaar krijg je dan een robuuster financieel stelsel.
Toekomst? 'Het mooist zou ik het vinden als we iets kunnen leren over hoe nieuwe landen, zoals Zuid-Soedan, beurzen zouden moeten opzetten', aldus Van Dijk.

Publicatie : Quest bijlage
Datum : 31 dec 2013
cm2 : 2.827
Pagina : 52
Advertentiewaarde : onbekend
Regio : Nederland
Frequentie : incidenteel
Oplage : 183.348

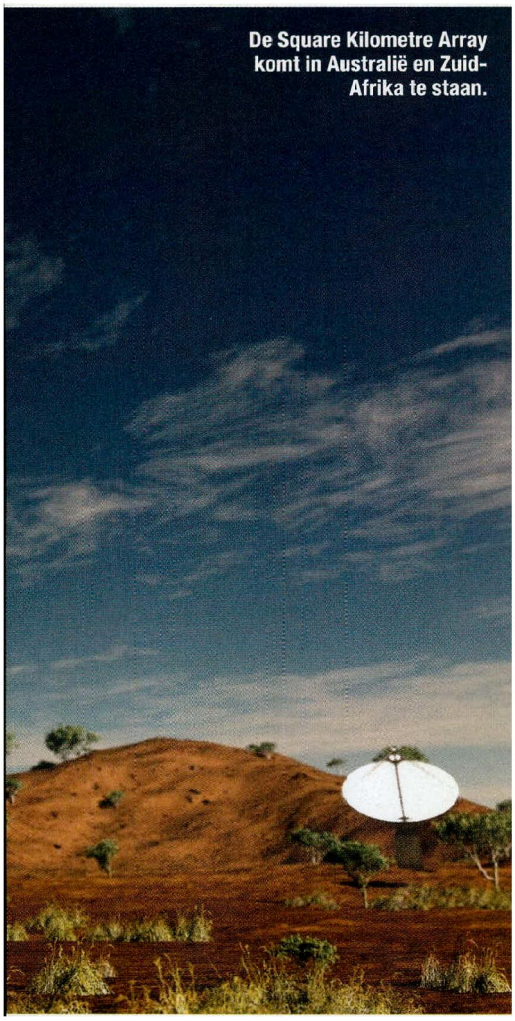
's Werelds grootste telescoop zal een schier onwerkbare berg data generen



Hoe verwerk je een miljard gigabyte per dag?

Wat is het? Zoektocht naar manieren om de verpletterende hoeveelheid gegevens uit een toekomstige radiotelescoop te verwerken.
Waarom bijzonder? In 2017 verrijst SKA, de grootste radiotelescoop aller tijden. Hij zal 1000 keer meer gegevens gaan produceren dan de grootste telescoop tot nu toe. En de data daarvan kunnen we al niet bijhouden.
Onderzoek? De sterrenkundigen zien het helemaal zitten, als de ingenieurs het maar mogelijk kunnen maken. SKA bestaat straks uit een combinatie van duizenden kleine schotels en tienduizenden tentstokachtige antennes. De ontvangers beslaan dan samen een vierkante kilometer (vandaar de naam SKA, Square Kilometre Array). De grootste opgaves voor de ingenieurs: hoe krijg je alle gegevens van die antennes naar één plek om ze te verwerken?
Hoe zorg je dat de centrale computer niet direct vastloopt? En hoe doe je dat een beetje milieuvriendelijk en goedkoop, want voor je het weet rijst het energieverbruik de pan uit. IBM-ingenieur Rik Jongerius, werkzaam bij ASTRON, het Nederlands instituut voor radioastronomie: 'We denken bijvoorbeeld na over hoeveel antennes in een groep bij elkaar moeten staan, en of we aan de rand van een veld antennes alvast een computer moeten neerzetten die het kaf van het koren scheidt. Dat scheelt in de hoeveelheid data die we door een glasvezelkabel moeten sturen naar de centrale supercomputer.'
Cijfer? Eén iPod per Nederlander. Zo veel gegevens gaat de supertelescoop SKA elke dag produceren. Het is ondoenlijk om al die

Publicatie : Quest bijlage
Datum : 31 dec 2013
cm2 : 2.827
Pagina : 52
Advertentiewaarde : onbekend
Regio : Nederland
Frequentie : incidenteel
Oplage : 183.348



De Square Kilometre Array komt in Australië en Zuid-Afrika te staan.

gegevens weken, maanden of jaren te stallen op harde schijven of servers. Daarom moeten de data in realtime verwerkt worden.
Nut? Als je de rekentijd halveert, verbruik je ook de helft van de energie. Dat scheelt al snel honderdduizenden euro's per jaar.
Toekomst? Bij gewone chips zorgen elektrische stroompjes voor de nullen en enen. De technici willen chips maken die met licht werken. Dat gaat veel sneller. Uiteindelijk komen de snelle lichtchips ook in je pc en in de servers van bedrijven en bij datacentra.

De beste tweelingen ter wereld

Wat is het? De koppeling van de DNA-gegevens van duizenden tweelingen met de vragenlijsten die ze al 25 jaar invullen.
Waarom bijzonder? Nederland heeft het beste tweelingenregister van de wereld. Daarmee kunnen we van allerlei zaken onderzoeken of ze door de omgeving veroorzaakt worden of dat ze erfelijk bepaald zijn. Denk aan rookverslaving, gezond eten, sporten en intelligentie.
Onderzoek? 'Mathijs, kun je het DNA van onze tweelingen vergelijken met het rookgedrag?' 'Ehm, Mathijs, wil je de database doorzoeken op verschillen in intelligentie tussen eeniige en tweeeiige tweelingen en tussen elkaar?' 'Mathijs, wanneer heb je de outputfile van mijn depressievraag binnen?' Dat zijn de vragen die Mathijs Kattenberg, programmeur bij het Tweelingenregister van de Vrije Universiteit Amsterdam, dagelijks kreeg. Het tweelingenregister was een grootverbruiker van BiG Grid. Kattenberg: 'Als je de complete DNA-volgorde van duizenden mensen in een database stopt, loopt de database vast en zie je door de bomen het bos niet meer. Wij werken daarom met markers. Dat zijn stukken DNA waarvan bekend is dat ze vaak tussen personen verschillen. Per persoon heb je het dan nog over een miljoen markers. Dat is nog steeds veel, maar het is 1000 keer minder dan als je het hele DNA zou bestuderen.'
Cijfer? 100.000 tweelingen zitten er in de database van de VU. Veel tweelingen beantwoorden al 25 jaar vragen over hun gewoontes, hobby's, verslavingen en voorkeuren. De laatste jaren komen er steeds meer gegevens over hun DNA bij en wordt de familie van de tweelingen gevolgd.
Nut? Dankzij het onderzoek met de tweelingen weten we nu bijvoorbeeld dat het voor het gedrag van een kind weinig uitmaakt of hij een meester of juf heeft. Dat helpt bij debatten in de Tweede Kamer over 'meer meesters voor de klas'. Ook weten we dat mannen sneller verslaafd raken aan roken en drank dan vrouwen en hoe dat ongeveer komt. En daarmee kunnen dokters en hulpverleners hun voordeel doen.
Toekomst? In de toekomst zullen mensen hun DNA laten onderzoeken om ziektes voor te zijn. Dat zal nieuwe problemen opleveren. Want waar bewaar je al die gegevens en moet iedereen wel alles willen weten?

Wie van hen steekt als eerste een peuk op? En kun je dat aflezen aan hun DNA?