



Too big data?

Er wordt veel verwacht van de mogelijkheden van big data en Internet of Things. Terecht, want het slim combineren van neutrale, real time data kan veel onnodige handelingen voorkomen. In veel scenario's lijken we er van uit te gaan dat deze groei ongelimiteerd kan doorgaan. Wat gebeurt er als dat niet het geval blijkt te zijn?

Data-obesitas

We gaan slordig om met onze data en veel partijen gaan er van uit dat we ongelimiteerd data kunnen toevoegen. Zo zie je systemen waar de data voor de zekerheid naar verschillende servers wordt gekopieerd om sneller analyses te kunnen uitvoeren. Als ik het principe van de Hadoop-technologie goed begrijp, worden willekeurige data op meerdere servers opgeslagen om te zien of een NoSQL-systeem hier zinvolle verbanden weet aan te tonen (laten we het maar registreren, want je weet nooit waar het goed voor is.)

De enorme groei aan data werkt verslavend en leidt tot een digitaal overgewicht. Jan Kees de Jager gaf recent aan dat van alle data vanaf het begin van de mensheid 90% is geproduceerd in de laatste 2 jaar! Lopen we tegen de grenzen aan van wat digitaal mogelijk is? In Londen kwamen op 11 mei 2015 telecombedrijven, topingenieurs en fysici bijeen om daarover te praten. Zie <http://www.welingelichtekringen.nl/tech/449405/stort-het-internet-in-overacht-jaar.html>

Is er geen eind aan de mondiale digitale geheugencapaciteit? Wachten we tot die grens bereikt is of sorteren we daar op voor door nu al te bedenken welke data we nu echt nodig hebben en welke niet?

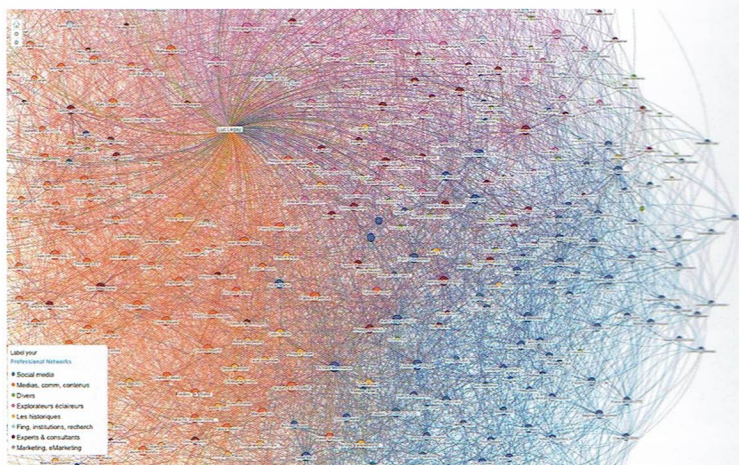
ROT en Dark data

Het Databerg Report 2015 stelt dat slechts 11% van de opgeslagen gegevens van bedrijven waarde heeft. Gemiddeld 32% betreft Redundant, Obsolete and Trivial (ROT) data (in Nederland 44%, Denemarken 48%, enz.). Daarnaast spreekt het rapport van data 'beneath the line of sight of senior management'; zogenaamde Dark data. Ook hier indrukwekkende cijfers: in Duitsland gaat het om 66%, in het Verenigd Koninkrijk om 58% van de data van bedrijven. In veel

gevallen gaat het om persoonlijke informatie: foto's vormen 57 % van de dark data, muziek 45%. (<http://www.emerice.nl/achtergrond/bedrijven-worstelen-databerg-infographic>)

Energiegebruik

Een bijkomend probleem is het enorme energiegebruik voor de opslag van die data. Een datacenter slaat zijn data op in zogenaamde racks, waarvan de hoeveelheid kan verschillen per m². Werd er eerst met 2kW per rack gerekend, tegenwoordig wordt al uitgegaan van 4,8 kW per rack, omdat de 'systemen zwaarder worden'. Het rapport 'Energiegebruik Nederlandse commerciële datacenters



2014-2017' van CE Delft berekent het energieverbruik van de Nederlandse datacenters in 2014 op 1,62 TWh per jaar.

Veiligheid

Er is een continue stroom van updates en criminelen of hackers die die updates proberen te omzeilen. Zo houden talloze mensen zich bezig met het beveiligen van de toegang tot, grotendeels nutteloze, data.

Datadiet

Wordt het geen tijd om na te gaan

denken hoe we onze dataconsumptie kunnen beperken, voordat we erdoor worden vermorzeld? We zullen waarschijnlijk moeten kiezen welke informatie we eigenlijk nodig hebben en welke niet per sé. Een belangrijke restrictie kan al worden bedacht voor de informatie die real time beschikbaar moet zijn en welke niet. Laten we alvast wennen aan het idee dat er waarschijnlijk een einde komt aan de digitale opslagcapaciteit en kijken wat we echt nodig hebben. Microsoft beperkt al de opslagcapaciteit voor gebruikers van Office 365, Personal en University tot 1 Tb i.p.v. onbeperkt. Voor de gratis versie gaat deze van 15 Gb naar 5 Gb.

ICT = goudmijn

Voor uitvinders is ICT een goudmijn: voor het sneller en complexer maken van systemen zijn veel vindingen nodig. Snellere, complexere systemen gebruiken

Big data: een onoverzichtelijke kluit van verbindingen

Bron: www.flickr.com

grosso modo meer energie, zodat er ook voor het verminderen van energieverbruik vindingen nodig zijn. Tenslotte zijn er vindingen nodig voor het beveiligen van data. Er ontstaat een vicieuze cirkel van sneller, groter en meer toegang die vraagt om zwaardere systemen met betere beveiliging: een goudmijn voor uitvinders.