

NORTH C INSIGHTS 20/01
DATA GRAVITY TRENDS IN NEDERLAND

De impact van data gravity op IT

12 tips voor
een succesvolle
digitale
transformatie



DUTCH
DATA CENTER
ASSOCIATION

NorthC
Datacenters

COLOFON

Het rapport “De impact van data gravity op IT” is een publicatie van Digital Gateway to Europe (DGTE) en Pb7 Research in samenwerking met Stichting Dutch Data Center Association (DDA) en NorthC Datacenters.

Editie

De impact van data gravity op IT,
Mei 2020

NorthC insights 20/01
Data Gravity trends in Nederland

Uitgave

NorthC Datacenters
Dutch Data Center Association
PB7 Research

Research

DGTE
PB7 Research

Artwork

Sixtyseven

Beschikbaarheid

Publicaties zijn te downloaden via
www.northcdatacenters.com

Contact

www.northcdatacenters.com

© 2020

Gebruiksvoorwaarden en disclaimer

“De impact van data gravity op IT” (hierna: ‘het Rapport’) bevat informatie en data samengesteld en/of verzameld door de DGTE (naar alle informatie en data wordt hierna verwezen als ‘Data’). De informatie in het Rapport is gebaseerd op bronnen die naar onze mening betrouwbaar zijn, maar DGTE staat niet in voor de nauwkeurigheid of volledigheid ervan. DGTE behoudt zich het recht voor op elk moment de inhoud aan te passen of onderdelen te verwijderen zonder daarover aan u mededeling te hoeven doen.

Hoewel DGTE zich inspanst om ervoor te zorgen dat de samengestelde/verzamelde Data nauwkeurig wordt weergegeven in het Rapport, verstrekt de DGTE de Data zoals deze beschikbaar is en zonder enige vorm van garantie met betrekking tot de inhoud of volledigheid. DGTE is nooit aansprakelijk voor enig gebruik of vertrouwen op de Data, inclusief, maar niet uitsluitend, voor enige interpretatie, beslissing of andere actie gebaseerd op de Data in het Rapport. Het kan zijn dat andere partijen belang hebben bij een deel van de Data in het Rapport.

DGTE staat er op geen enkele manier voor in of garandeert dat het eigendom en controle heeft op alle rechten met betrekking tot de Data en de DGTE is niet aansprakelijk tegenover gebruikers voor claims tegen gebruikers door derden in verband met het gebruik van enige Data.

DGTE en zijn werknemers staan op geen enkele manier garant en bekrachtigen op geen enkele manier de producten of services van derden gebaseerd op de Data, het materiaal of de inhoud/verwijzingen naar inhoud van het Rapport.

INHOUDSOPGAVE

Inleiding	6
 1. WAT IS DATA GRAVITY EN WAAROM IS HET BELANGRIJK?	
Wat is Data Gravity	9
De data-explosie	10
Samenvattend	12
 2. DATA GRAVITY EN DE CLOUD	
De aantrekkingskracht van de cloud	14
Wat doen we met de silo's?	16
De contouren van een data-first strategie	17
Samenvattend	18
 3. EDGE GRAVITY	
Edge computing	20
Drijfveren	21
Sectoren en use-cases	22
Hoe past de edge in het datacenter landschap?	24
Samenvattend	26

INHOUDSOPGAVE

4. ROL VAN HET REGIONALE DATACENTER

Rol van het regionale datacenter	28
Cloud en het datacenter	29
Het datacenter als knooppunt	30
Datacenters en de edge	30
Samenvattend	32

5. ANALYSE EN AANBEVELINGEN

Analyse	34
Is door data gravity gedistribueerde computing weer helemaal terug?	36
Aanbevelingen - Enterprise computing	37
Aanbevelingen - Edge computing	38
Ten slotte	39

APPENDIX

Methodologie	41
Definities	42
Bronnen	45

VOORWOORD

Ons leven en onze economie worden steeds digitaler en we zitten daarbij volop in een transformatie. Data krijgen een steeds centralere rol en steeds meer gewicht in tal van processen en beslissingen om bedrijven effectiever te laten concurreren. Deze transformatie gaat in een snel tempo en kent vele uitdagingen. Om in een steeds digitaler wordende wereld te blijven overleven, willen bedrijven steeds meer en sneller data analyseren, in een tijd dat de datastromen alsmaar groter worden en uit steeds meer bronnen tegelijkertijd komen.

Beheersbaarheid van data en IT was vroeger een relatief eenvoudig proces, toen organisaties al hun gegevens centraal in on-premises IT-omgevingen opsloegen. In een cloudgefocuste wereld en in een wereld van het 'Internet of Things' is die beheersbaarheid veel gecompliceerder en dat vraagt om een andere aanpak en keuzes. Daarnaast is het juist nu van groot belang om ons aan de plots veranderende omstandigheden aan te passen. In een tijd waar eisen aan continuïteit, gegevensbescherming en bereikbaarheid van diensten steeds hoger worden.

Data dicteren ook steeds meer de keuzes die bedrijven moeten maken over waar applicaties en services zich binnen de infrastructuur moeten bevinden. Dit rapport legt uit en analyseert wat de impact is van de steeds groter wordende kracht van data, ook wel 'data gravity' genoemd, op de bedrijfsprocessen. Daarnaast fungeert dit rapport als handleiding voor de stappen die bedrijven moeten nemen om op een goede manier met de toenemende datastroom om te gaan. Met dit rapport willen we organisaties beter inzicht geven in het fenomeen data gravity en aangeven welke strategieën zij kunnen volgen.

Het rapport bevat bovendien ook twaaalf tips die organisaties kunnen helpen bij hun digitale transformatie.

NorthC Datacenters

INLEIDING

De hoeveelheid data die we creëren groeit exponentieel. En we zien steeds meer kansen om met die data waarde te creëren. Daarom bouwen we steeds meer toepassingen die waarde uit de data halen, analyseren en interpreteren. Met behulp van data en data-analyse tools kunnen we slimmere beslissingen nemen binnen een organisatie. Niet alleen op strategisch niveau, maar juist ook om operationele taken en de business te ondersteunen of te automatiseren. Terwijl de hoeveelheid data die we willen analyseren steeds groter wordt, moeten die data tegelijkertijd steeds sneller daar beschikbaar zijn waar deze nodig zijn. Vertraging in de beschikbaarheid van data is niet meer acceptabel.

Data lijken eenvoudig te verplaatsen, maar dat is in de praktijk niet het geval. Grote hoeveelheden data transporteren is kostbaar en tijdsintensief. Ook het verplaatsen van een meer bescheiden hoeveelheid data kan ongewenst zijn. Zeker als de verwerking nodig is voor real-time handelingen op locatie. Omdat data steeds moeilijker van hun plek te krijgen zijn, en we dat ook steeds minder vaak willen, trekken de data applicaties en services als het ware steeds dichterbij naar zich toe. Dit fenomeen wordt wel 'data gravity' genoemd (McCrory, 2010).

Organisaties zullen hun IT-architectuur hierop moeten aanpassen, waarbij de datalocatie leidend wordt. Soms zal dat een centrale locatie zijn, maar steeds vaker zal dat dichtbij het eindpunt zijn. Dat eindpunt kan een slim apparaat zijn, maar ook een klant of een andere gebruiker.

In dit rapport onderzoekt Pb7 Research samen met de Dutch Data Center Association en NorthC Datacenters de opkomst en ontwikkeling van data gravity, de impact daarvan op het Nederlandse IT-landschap en hoe organisaties daar optimaal op kunnen inspelen. De hypothese die we daarbij toetsen, is:

'Door data gravity is gedistribueerde computing¹ weer helemaal terug'

¹ Gedistribueerde computing: bij gedistribueerde computing wordt een enkele taak over meerdere systemen opgesplitst in meerdere subtaken die relatief onafhankelijk berekend en achteraf weer samengevoegd kunnen worden tot een enkel resultaat.

Om de hypothese te toetsen, trachten we met behulp van deskresearch de volgende vragen te beantwoorden:

- In welke mate groeit de databerg?
- In welke mate worden toepassingen naar data toe getrokken?
- Waar bevinden zich onze data, hoe verandert dat onder invloed van data gravity en wat is de impact daarvan op enterprise computing en de cloud?
- Wat betekent data gravity voor de IT-architectuur en hoe houd ik dat beheersbaar?
- Wat is de rol van nationale datacenters met een regionale dekking?

Leeswijzer

In dit rapport gaan we eerst in op wat data gravity is en waarom het belangrijk is. In het tweede hoofdstuk gaan we in op welke veranderingen dat teweegbrengt en zou moeten brengen in enterprise en cloud computing. Het derde hoofdstuk gaat dieper in op hoe de snelle opkomst van data gravity aan de rand van het netwerk vraagt om een nieuw gedistribueerd datacenter landschap. En in het vierde hoofdstuk gaan we in op de mogelijke rol die daarin is weggelegd voor nationale datacenters met een regionale dekking. Na de analyse waarin we de hypothese toetsen aan de hand van bovenstaande vragen, eindigen we met een set concrete aanbevelingen.

WAT IS DATA GRAVITY EN WAAROM IS HET BELANGRIJK?



DATA GRAVITY

Data gravity is een term die bijna tien jaar geleden voor het eerst werd gebruikt. Het werd al snel opgepikt als een interessante en relevante notie, maar werd nog niet breed gedragen. De laatste tijd krijgt het steeds meer aandacht en groeit de urgentie om er wat mee te doen. In dit hoofdstuk gaan we in op wat er wordt bedoeld met data gravity en leggen we uit waarom het alsmaar belangrijker wordt.

Wat is data gravity?

Het zo dicht mogelijk bij elkaar houden van data en de bijbehorende applicaties en services heeft een groot voordeel. Het zorgt namelijk voor een zo hoog mogelijke doorvoersnelheid. Omdat data het minst eenvoudig van hun plek te krijgen zijn, brengen we applicaties en services steeds vaker naar de data toe. Elke concentratie van data, klein of groot, trekt dus applicaties en services als het ware naar zich toe. Dat fenomeen heet 'data gravity'.

Voor bedrijven manifesteert die 'zwaartekracht' zich op diverse locaties. Vergelijk het met ons zonnestelsel: de zon is bijvoorbeeld 'de' centrale publieke cloud², datacenters³ zijn de planeten en locaties en devices zijn de manen en satellieten. Het reizen tussen objecten is kostbaar. Het kost veel tijd en brengt risico's met zich mee.

Hoe meer data naar de cloud (de zon) gaan, hoe gewenster het is om data-analyse tools en allerlei centrale applicaties naar de zon te brengen. Maar als je de uitkomsten heel snel wilt hebben op je hoofdkantoor, dan is de reis van de zon naar de aarde zelfs op lichtsnelheid eigenlijk te lang. En als je met behulp van data-analyse real-time een machine wil aansturen, of een zelfrijdende auto, of een gebruiksvriendelijke toepassing op een smartphone, dan is die planeet (het datacenter) zelfs te ver weg.

² Cloud: via internet op aanvraag toegang tot een gedeelde verzameling eenvoudig aanpasbare computermiddelen zoals servers, software en dataopslag. Diensten worden naar behoefte afgenomen zonder hulp van de aanbieder (selfservice), betalen is op basis van werkelijk gebruik.

³ Datacenter: een geconditioneerde ruimte waarin servers en bijbehorende apparatuur worden geplaatst. Ook serverruimtes binnen gebouwen met andere functies rekenen we in dit document tot datacenters.

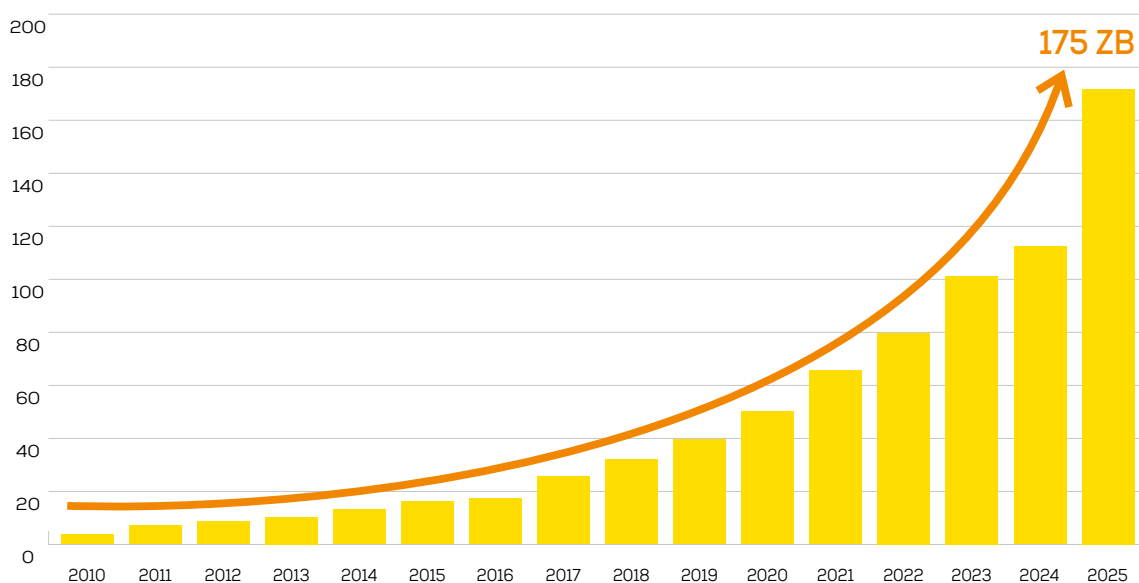
Vandaar dat we onze smartphones vol met apps zetten en dat we steeds meer rekenkracht aan de rand van het netwerk plaatsen.

Naast de grote hoeveelheid datalocaties waar we tegenwoordig rekening mee moeten houden, zien we ook dat de hoeveelheid data heel snel groeit. De aantrekkende kracht neemt daardoor logischerwijs nog verder toe, maar ook de complexiteit van het data gravity landschap.

De data-explosie

De hoeveelheid data die we met elkaar creëren en gebruiken, groeit exponentieel. Onderzoeksbureau IDC schat dat het wereldwijde datavolume in de komende jaren met gemiddeld 27% per jaar groeit (IDC, 2018). Dat is een verdubbeling elke drie jaar. Dankzij de nog altijd snel toenemende kracht van processors, kunnen we die data produceren, verwerken en er steeds meer toepassingen voor bedenken. En door de snel toenemende bandbreedte en snelheid van vaste en mobiele netwerken, kunnen we deze data best snel verplaatsen. Een willekeurige pagina van een website verschijnt binnen een seconde op ons scherm en we streamen zonder problemen een HD-film. Dat deze af en toe buffert, nemen we maar voor lief. Het zijn kleine vertragingen die we accepteren en die we bijna niet meer zien. Eigenlijk zie je pas echt goed hoe moeilijk data te verplaatsen is als je heel erg veel data in een keer wil verplaatsen. Het is meestal een stuk sneller om een volle harde schijf handmatig te verplaatsen, dan de bestanden die erop staan via een netwerk te versturen. Voor veel toepassingen kan zelfs een minieme vertraging een groot probleem zijn. De data, de verwerking en de gebruiker hou je dan het liefst heel dicht bij elkaar.

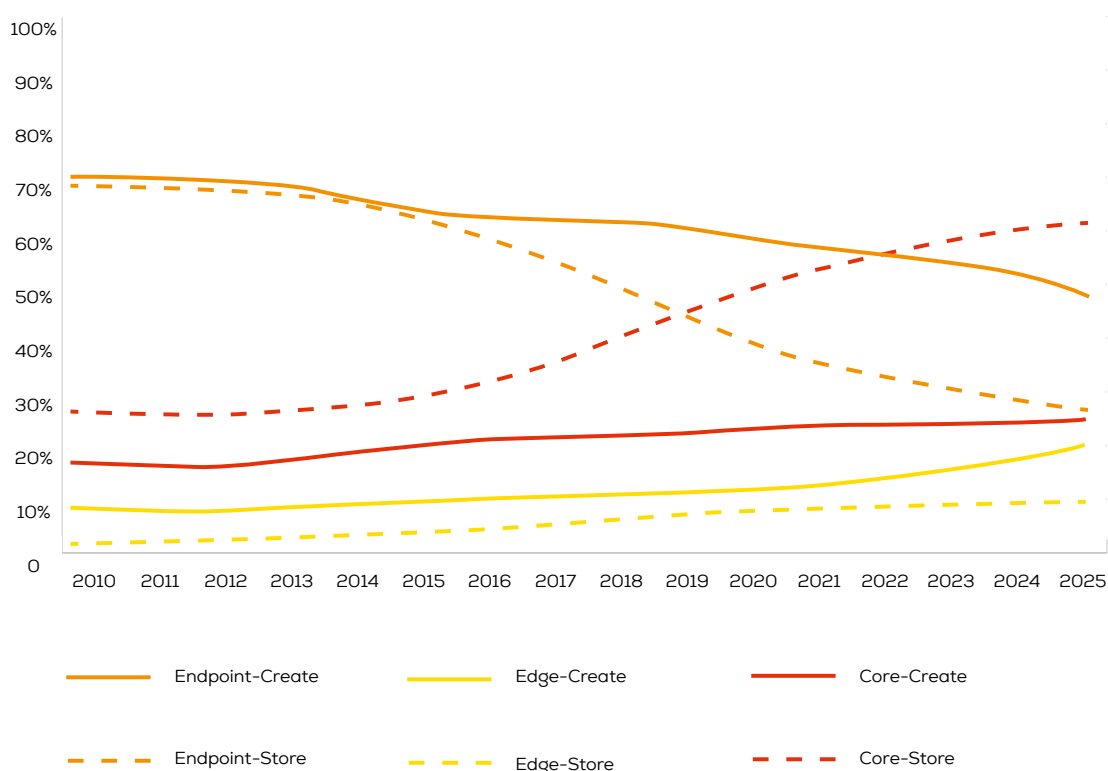
Annual Size of the Global Dataphere



Bron: IDC, 2018

Kijken we naar het data gravity landschap, dan zien we dat we door de opkomst van de cloud momenteel met een sterke centralisatietrend te maken hebben. Volgens IDC werd rond 2010 meer dan twee derde van alle data door 'endpoints' (eindpunten), zoals PC's en gaandeweg ook smartphones en tablets, gegenereerd en werd het daar ook opgeslagen. Tegenwoordig zien we dat de data steeds meer in centrale locaties worden opgeslagen, zoals bedrijfsdatacenters en de cloud. De cloud heeft daarbij steeds vaker de voorkeur. Daarnaast zien we ook nog dat steeds meer data gegenereerd en opgeslagen worden ergens tussen het eindpunt en die centrale locatie, aan de randen van het netwerk, oftewel de 'edge'⁴. Als de data bij de edge worden verwerkt, hebben we het over edge computing⁵.

Creating and Storing Data by Core/Edge/Endpoint



Bron: IDC, 2018

⁴ Edge: de rand van het netwerk; wat niet tot de kern behoort.

⁵ Edge computing: verwerking van data aan de rand van het netwerk, dichtbij (maar niet op) het apparaat.

Samenvattend

Data trekken applicaties en services naar zich toe. Dit fenomeen noemen we data gravity. Computing vindt steeds meer plaats op de locatie waar de meeste waarde kan worden geleverd. Dat is vaak op een centrale plek, maar ook steeds vaker aan de rand van het netwerk. Door de snelle groei van data en toepassingen wordt data gravity steeds tastbaarder. Bedrijven zullen daarom hun IT-strategie daarop moeten afstemmen. In de volgende hoofdstukken zoomen we hier meer op in. Eerst gaan we in op de impact van data gravity op de cloudstrategie. Daarna komt edge computing ruimschoots aan bod.

DATA GRAVITY EN DE CLOUD



DATA GRAVITY EN DE CLOUD

De term data gravity is in eerste instantie bedacht om de aanzuigende werking van de cloud te verklaren. In de cloud zien we een toenemende concentratie van data, applicaties en services. Doordat bijna alles naar de cloud gaat, lijkt het of datasilo's⁶ worden afgebroken en verruild voor een krachtige centralisatie. In dit hoofdstuk laten we zien dat daar een kern van waarheid in zit, maar dat de cloud ook juist tot een meer gedistribueerd landschap leidt.

De aantrekkingskracht van de cloud

Organisaties verplaatsen steeds meer applicaties naar de cloud, waardoor ook steeds meer data naar de cloud verschuiven. In Nederland volgen organisaties daarbij voornamelijk het tempo van applicatievernieuwing. Dat betekent als een applicatie aan vervanging toe is, kiest men indien mogelijk een SaaS⁷- of een andere cloudoplossing; als nieuwe functionaliteit nodig is, komt ook eerst de cloud in beeld; en als er een nieuwe applicatie ontwikkeld dient te worden, gebeurt dat eveneens in de cloud.

Deze 'cloud-first' trein dendert in Nederland volop door in de kritieke bedrijfs-IT, ofwel 'enterprise computing'. De trein wordt voortgedreven door de belofte van kosteneffectieve, snelle en schaalbare toegang tot capaciteit (compute en store) en functionaliteit. Het vuur wordt verder opgestookt door een stevige drang van de softwareleveranciers die juist vol op SaaS inzetten. Het applicatie voor applicatie overstappen naar de cloud heeft een aantal duidelijke voordelen. Zo biedt het ruimte om de meest geschikte cloudoplossing te vinden voor individuele applicaties. Maar het belangrijkste is wel dat daarmee de overstap naar de cloud hanteerbaar blijft. Door de overstap in heel veel kleinere stappen te doen, ontstaat er tijd en ruimte om fouten te maken en te leren.

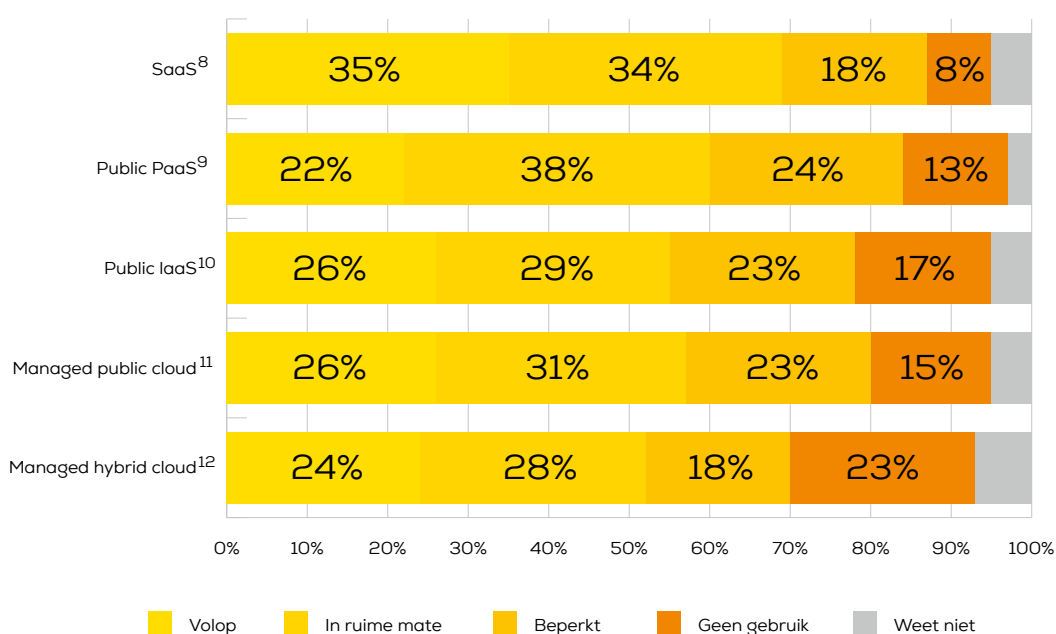
⁶ Datasilo: een term om aan te geven dat veel data in aparte databases zitten opgesloten.

⁷ SaaS: clouddienst waarbij een applicatie als dienst wordt afgenomen.

Het grote risico van deze cloud-first aanpak is dat het grote geheel uit het zicht dreigt te verdwijnen. Nederlandse organisaties hebben de laatste vijf tot tien jaar veel ruimte geboden om te experimenteren met cloud. Een gevolg daarvan is dat business managers, met of zonder hulp van de IT-afdeling, zelf naar cloudoplossingen zijn gaan zoeken. Daardoor is een kleurrijk en vaak complex IT-landschap ontstaan binnen bedrijven. En de bijbehorende data zitten nu opgesloten in silo's bij een groot aantal cloudaanbieders.

Van welke van de volgende vormen van cloud maakt uw organisatie gebruik?

N=200 (Nederlandse organisaties met 100 of meer medewerkers).



Bron: IDG, 2019

⁸ SaaS: clouddienst waarbij een applicatie als dienst wordt afgenomen.

⁹ PaaS: clouddienst waarbij een platform als dienst wordt afgenomen en waar applicaties op draaien en op maat ontwikkeld.

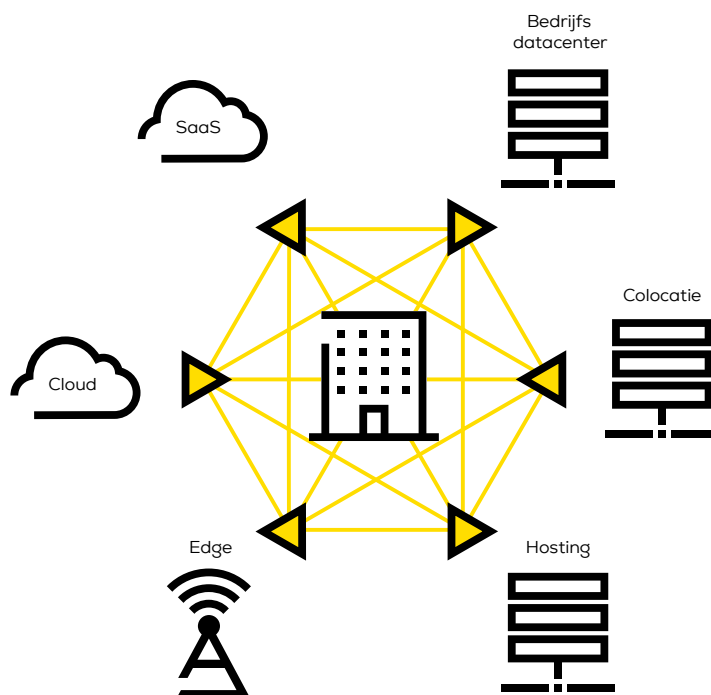
¹⁰ IaaS: clouddienst waarbij IT-infrastructuur, zoals opslag en verwerkingscapaciteit als dienst wordt afgenomen.

¹¹ Managed cloud: dienst waarbij een service provider afzonderlijke publieke en/of private cloudmiddelen als een geheel inricht en beheert.

¹² Hybride cloud: cloudomgeving waar een private cloud workload kan worden geschaald naar een publieke cloud, eventueel vice versa.

Wat doen we met de silo's?

Doordat veel bedrijven de afgelopen jaren vooral per workload¹³ hebben bepaald welke cloudoplossing de juiste is, zijn er veel nieuwe silo's gecreëerd. Zelfs als de IT-afdeling het voor elkaar heeft dat er gestandaardiseerd wordt op een enkel public cloudplatform¹⁴, dan wordt nog altijd de ene na de andere applicatie als SaaS-dienst of in de vorm van hosting afgenomen. Een aanzienlijk deel van de workloads blijft ook gewoon on-premises¹⁵, bijvoorbeeld in een private cloudomgeving.¹⁶



Figuur: Zelfs in een sterk vereenvoudigd schema als hierboven is duidelijk dat data tegenwoordig verspreid zijn over een grote diversiteit aan locaties.

Doordat data opgesloten blijven in silo's, ontstaat er een aantal problemen. Niemand heeft goed zicht op waar alle data zijn. Als een klant in het kader van de AVG bijvoorbeeld vraagt welke persoonlijke gegevens er over hem of haar worden bewaard, dan hebben de meeste organisaties een groot probleem. Het kost ze bijzonder veel moeite om dat te achterhalen en men kan niet garanderen dat er niet nog meer gegevens ergens opgeslagen zijn.

¹³ Workload: een hoeveelheid rekenwerk; afzonderlijke rekentaak.

¹⁴ Public cloud: clouddiensten waarbij de middelen door meerdere afnemers (uit verschillende organisaties) worden gedeeld.

¹⁵ On-premises: op een locatie van de gebruiker (bedrijf of instelling).

¹⁶ Private cloud: clouddiensten waarbij middelen door verschillende interne klanten (afdelingen, dochters en/of applicaties) worden gedeeld.

Een ander probleem is dat data, mede door data gravity, moeilijk uit de silo's zijn te krijgen. Er zijn weliswaar steeds meer mogelijkheden om data over meerdere databases heen te kunnen analyseren, maar iedere oplossing heeft zijn eigen voor- en nadelen. Wat blijft, is dat de fysieke afstand tussen de databases voor ongewenste vertraging zorgt.

Een volgend probleem van al die silo's ligt op het beheervlak in brede zin. Zo heeft IT-security vaak geen goed en volledig overzicht van welke data zich waar bevinden. Daardoor worden data uit verschillende databases niet altijd volgens dezelfde richtlijnen beschermd. Ook zien we dat niet iedere applicatie aan dezelfde kwaliteits- en continuïteitseisen voldoet en dat ook het zicht daarop verre van compleet is. Van iedere applicatie en dataset zou een consistente risico-analyse beschikbaar moeten zijn die kijkt naar security, compliance en continuïteit.

Veel Nederlandse organisaties lopen tegen dergelijke problemen aan. Ze zetten daarom steeds meer in op een heldere, centrale cloudstrategie. Bovendien richten ze een governance organisatie in die ervoor moet zorgen dat cloud- en applicatieprojecten volgens afgesproken richtlijnen worden geïmplementeerd en beheerd. In de meeste gevallen richt deze aanpak zich nog altijd meer op losstaande projecten en applicaties dan op de data als geheel. Daardoor blijven veel silo's voorlopig nog in stand, maar zit er in ieder geval een rem op het ontstaan van nieuwe silo's. Het fundamentele probleem bij de aanpak hiervan is dat de Chief Information Officer (CIO) over de IT gaat, terwijl de business managers over de data gaan. En beiden zijn vaak best tevreden met hun eigen silo('s).

De contouren van een data-first strategie

De laatste twee jaar zien we dat veel organisaties een dimensie toevoegen aan hun IT-strategie: ze stellen data centraal en zoeken naar kansen om hier meer waarde uit te halen. De snelle opkomst van nieuwe mogelijkheden op het gebied van kunstmatige intelligentie¹⁷, machine learning¹⁸ en geavanceerde analytics helpt hen daarbij. De business cases zijn vaak bijzonder aansprekend. Bedrijven merken dat ze hiermee bedrijfsprocessen en de opbrengsten daarvan aanzienlijk kunnen optimaliseren.

Om van die mogelijkheden gebruik te maken, is toegang tot zoveel mogelijk data nodig, uit zoveel mogelijk systemen. En als dat niet genoeg is, worden extra sensoren geplaatst om nog meer data te verzamelen.

¹⁷ Kunstmatige intelligentie of artificial intelligence (AI) : het vermogen van een systeem om externe gegevens correct te interpreteren, om te leren van deze gegevens, en om deze lessen te gebruiken om specifieke doelen en taken te verwezenlijken via flexibele aanpassing.

¹⁸ Machine learning: het onderdeel binnen kunstmatige intelligentie dat zich bezighoudt met de ontwikkeling van algoritmes en technieken waarmee computers kunnen leren.

Om data uit verschillende systemen gecombineerd te kunnen analyseren, worden data steeds vaker op een centrale plek verzameld, met behulp van gefedereerde databases¹⁹, datahubs of ongestructureerde data lakes²⁰. In een data lake bijvoorbeeld worden data bevrijd uit starre, gestructureerde databases en datawarehouses. Ongestructureerde data kunnen naar hartenlust worden gecombineerd om tot nieuwe inzichten te komen.

Daarmee kunnen bijvoorbeeld veel diepere klantinzichten worden verworven of nieuwe genezingsmethoden worden gezocht.

Aangezien de data zich in toenemende mate in de cloud bevinden, ook als bedrijven niet van een datahub²¹ of data lake gebruikmaken, plaatsen gebruikers juist daar de tools om die data te analyseren, want voor veel bewerkingen kost het verplaatsen van data uit de cloud naar een eigen server te veel tijd. Bovendien brengt een cloud provider relatief veel kosten in rekening voor het downloaden van data uit de cloud.

Samenvattend

Enterprise computing ziet er onder invloed van de cloud anders uit dan voorheen. De aanzuigende werking van de cloud leidt niet per se tot een centralisatie van computing. Bedrijven hebben data niet alleen, net als voorheen, over veel databases verspreid, ze zijn nu bovendien over veel meer fysieke locaties verspreid. Daardoor dreigen ze de controle over de data kwijt te raken. Doordat Nederlandse organisaties steeds meer mogelijkheden herkennen om waarde uit data te halen, neemt de dataverzameling toe. Bedrijven starten nieuwe initiatieven om data uit de zwaartekracht van de silo's te ontsluiten. Daarbij stuiten ze vooral op de nodige uitdagingen. Het is van groot belang dat organisaties daar beter grip op krijgen, liefst voordat edge computing de complexiteit verder doet toenemen.

¹⁹ Gefedereerde database: virtuele database van waaruit query's kunnen worden gedraaid die direct worden opgehaald uit live databases.

²⁰ Data lake: verzamelplaats van data (kopie) uit een groot aantal bronnen waarbij de data ongestructureerd worden opgeslagen; typisch voor data science mogelijkheden.

²¹ Datahub: als een data lake, maar dan met deels of geheel geharmoniseerde data, indexering en standaard analytische tools.

3

EDGE GRAVITY



EDGE GRAVITY

Terwijl Nederlandse organisaties op zoek zijn naar de beste manier om de enterprise en cloud computing te organiseren, krijgen ze te maken met een heel nieuwe uitdaging: edge computing. Aan de rand van het netwerk laat data gravity zich in toenemende mate gelden. In dit hoofdstuk gaan we eerst in op edge computing zelf: wat is het en waarom is het belangrijk? Vervolgens gaan we in op hoe data gravity het datacenter landschap verandert.

Edge computing

Edge computing is de oplossing voor een groeiend data gravity probleem. Met behulp van goedkope sensoren, steeds betere vaste en mobiele netwerken en een explosie aan analytische software en kunstmatige intelligentie bedenken we in snel tempo nieuwe toepassingen die we willen gebruiken op plekken die ver weg zijn van centrale datacenters of de cloud. Dat noemen we edge computing. Digitale media en vooral video streaming zijn daarbij de absolute grootverbruikers die met name de consument bedienen. In de zakelijke markt zien we een explosie van Internet of Things²² (IoT-) oplossingen. Deze kunnen een grote rol gaan spelen in de digitale transformatie van een groot aantal sectoren. Voor een belangrijk deel van deze IoT-oplossingen is het onlogisch en onwenselijk om gebruik te maken van dataverwerking op een centrale locatie.

²² Internet of Things (IoT): een voorgestelde ontwikkeling van het internet, waarbij alledaagse voorwerpen zijn verbonden met het netwerk en gegevens kunnen uitwisselen.

Drijfveren

De voornaamste redenen om computing in centrale locaties te omzeilen en voor een andere oplossing te kiezen zijn:

- **Lage latency**²³: transport van data kost tijd. Kostbare tijd die voor bepaalde activiteiten niet acceptabel is. Door de opkomst van glasvezel en 5G-connectiviteit hebben we het vaak over niet meer dan enkele milliseconden. Als we die data naar de andere kant van de Atlantische Oceaan sturen, komen daar al snel 60 milliseconden bij. Voor sommige machine-to-machine oplossingen is dit simpelweg te veel tijd.
- **Netwerkkosten**: steeds meer apparaten verzamelen aanhoudend meer data. Een van de meest data-intensieve apparaten uit de IoT-stal zijn HD-beveiligingscamera's. Het permanent doorsturen van HD-beelden van een groot aantal camera's vraagt zeer veel bandbreedte. Dat betekent hoge investeringen in eigen netwerken en/of het huren van veel netwerkcapaciteit. Die kosten kunnen worden beheerst door het verkeer zelf te beperken tot data waarvan het transport strikt noodzakelijk is. Data kunnen immers ook lokaal worden opgeslagen, bewerkt en doorgestuurd. Grote contentaanbieders maken van dit principe volop gebruik. Als Netflix een populaire film in Nederland aanbiedt, sturen zij deze maar één keer digitaal de oceaan over. Nederlandse consumenten streamen de film naar hartenlust vanuit een Amsterdams datacenter.
- **Wet- en regelgeving**: het is niet altijd wenselijk of juridisch mogelijk om bepaalde data buiten de landsgrenzen of buiten de Europese Unie op te slaan of te verwerken. Om toch gebruik te kunnen maken van de clouddiensten van bijvoorbeeld Amerikaanse providers, maken deze partijen het mogelijk om data binnen Nederland of de EU te houden.

Bovenstaande problemen kunnen worden opgelost door de dataverwerking dichterbij het data-producerende eindpunt toe te brengen. Daar zijn verschillende mogelijkheden voor. Allereerst kunnen de data op het apparaat zelf verwerkt worden. Verwerking van data op het eindpunt noemen we 'on-device' of 'endpoint computing'²⁴. Het is echter niet altijd zinvol en/of economisch verantwoord om data op een apparaat zelf op te slaan of te bewerken. Bij een auto is dat vanzelfsprekend. Hebben we het echter bijvoorbeeld over een industriële IoT-oplossing, zoals een sensor in een klep op een willekeurige leiding of een sensornetwerk, dan is verwerking van data op het eindpunt niet kosteneffectief. Dit kan worden opgelost door de data te versturen naar een dichtbijgelegen locatie waar de verwerking plaatsvindt, waarna het resultaat wordt terug- of doorgestuurd naar het eindpunt om een handeling in gang te zetten. Die verwerking dicht bij het apparaat noemen we edge computing.

²³ Latency: de vertraging in de dataoverdracht over een datacommunicatienetwerk. Lage (of low) latency is een minimale vertraging.

²⁴ On-device of endpoint computing: verwerking van data op het apparaat zelf.

Sectoren en use-cases

Edge computing is dus vooral zinvol bij Internet of Things oplossingen die normaal gesproken veel bandbreedte vergen en/of real-time reactievermogen vragen en waar dataopslag of -verwerking op het apparaat ongewenst is. Dat maakt dat edge computing een belangrijke rol kan gaan spelen bij een breed scala aan digitale oplossingen die aan de basis staan van de digitale transformatie van organisaties binnen een groot aantal sectoren.

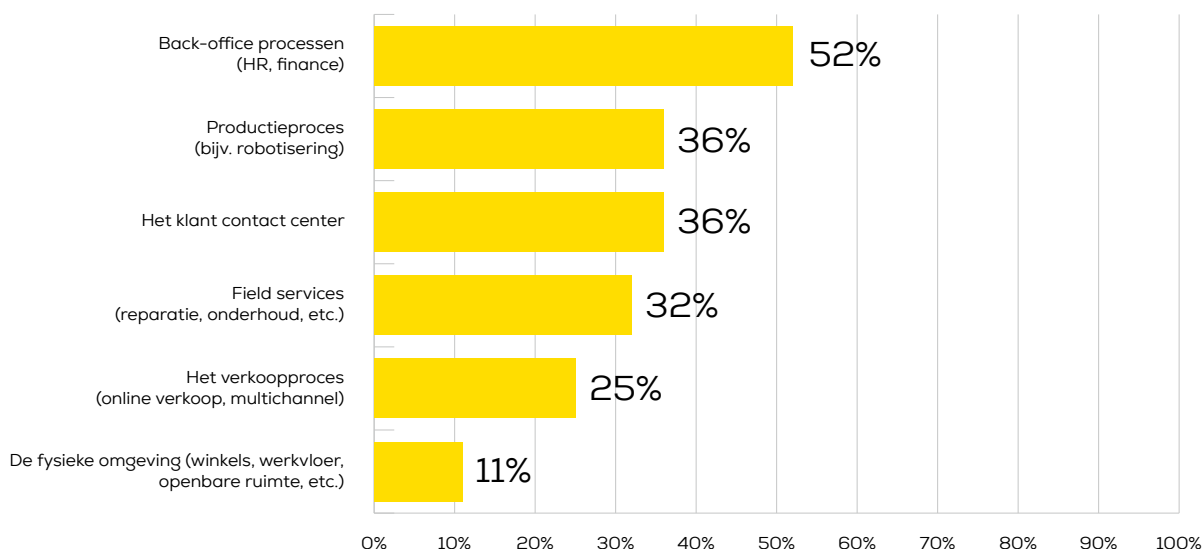
Voorbeelden van use-cases voor edge computing

Landbouw	• Precisie landbouw: irrigatie en bemesting op de vierkante meter • Stappentellers voor vee
Industrie	• Sensornetwerken voor performance optimalisatie, onderhoud, kwaliteitscontrole • Robotisering op de fabrieksvloer
Detailhandel	• Dynamische in-store pricing • 3D/4D printen • AR/VR
Finance	• Asset tracking (lease vloot) • Slimme geldautomaten
Transport en logistiek	• Autonome voertuigen • Robotisering en distributiecentra • Transparante supply chain (tracken en documentatie)
Overheid	• Verkeersstromen en parkeren • Openbare veiligheid (gezichtsherkenning, crowd control) • Beheer wegen en waterwerken
Energie	• Smart grid • Smart metering
Zorg	• Domotica • Mobiele monitoring
Overig	• Beveiligingscamera's • Mobile gaming • Smart buildings • Smart home

In de industriesector spreken we bijvoorbeeld over industrie 4.0, een nieuwe automatiseringsslag gebaseerd op slimme robotisering. Duitse producenten zetten daar al vol op in, terwijl veel Nederlandse fabrikanten er de komende jaren nog mee aan de slag moeten. De verwachting is dat de industrie, samen met de energiesector, uitgroeien tot de belangrijkste afnemers van edge computing. Maar er zijn ook veel kansen in andere sectoren: van de landbouw tot de zorg, en van het bankwezen tot Rijkswaterstaat. Overal zien we bedrijven, publieke sectoren en consumenten investeren in digitale transformatie aan de edge.

In welke van de volgende vormen van digitale transformatie worden er binnen uw organisatie de komende jaren geïnvesteerd? Transformatie van...

(N=200, Nederlandse organisaties met 100 of meer medewerkers).



Bron: IDG, 2019

In totaal verwacht onderzoeksbureau Pb7 Research dat Nederlandse bedrijven binnen vijf jaar meer dan 60 miljoen euro per jaar aan edge computing uitgeven. In dat cijfer zitten de uitgaven aan hardware, zoals edge nodes²⁵/gateways, sensoren en routers, software, managed edge platforms en diensten. Het gaat hierbij om de uitgaven die daadwerkelijk ultralokaal zijn, waarbij we kijken naar een combinatie van private edge²⁶, publieke oplossingen op het niveau van microdatacenters²⁷ en ultra lokale cloudoplossingen.

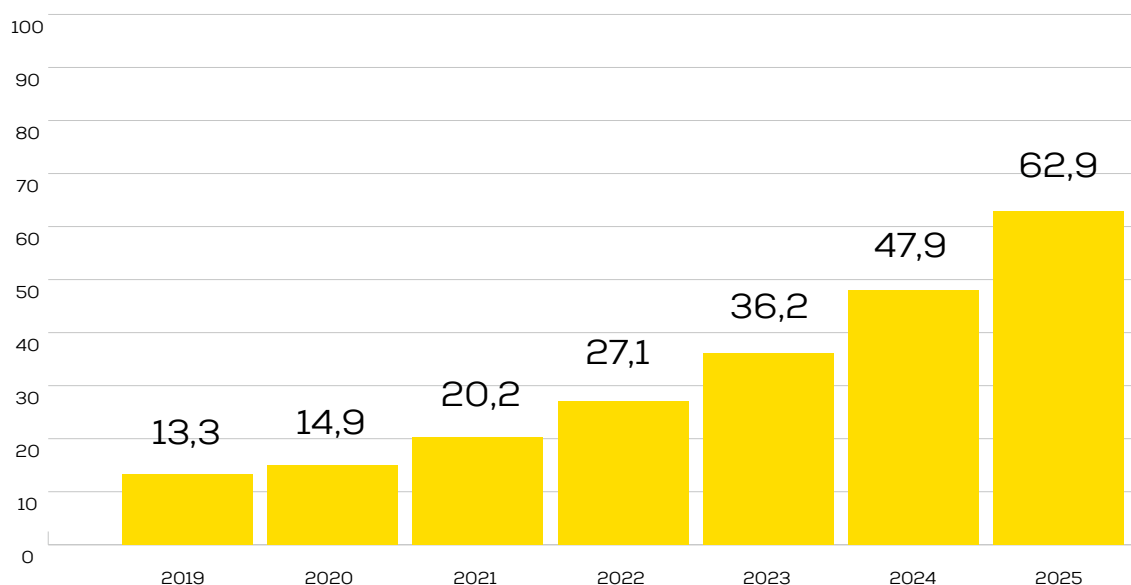
²⁵ Edge nodes (ook wel edge gateways genoemd): computers die in de edge data deels of geheel verwerken voor endpoints (apparaten) en bovendien vaak als een cluster kunnen fungeren.

²⁶ Private edge: edge oplossingen die gehuisd worden op een bedrijfseigen locatie.

²⁷ Microdatacenter: zeer klein, autonoom functionerend datacenter dat op zeer kleine afstand functioneert van endpoint devices (<10 km).

Uitgaven (EUR Mln.) van Nederlandse bedrijven aan Edge Computing, 2019-2025

CAGR 2020 tot 2025 = 33%



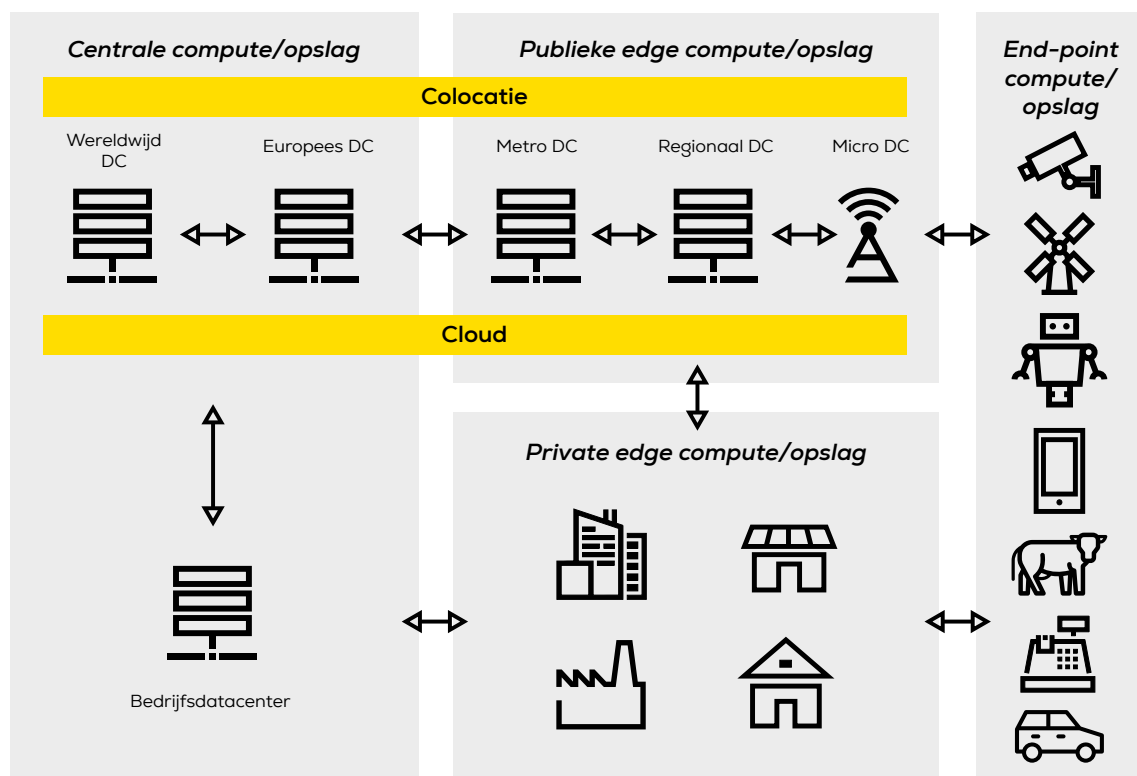
Bron: Pb7 Research, 2019

Hoe past de edge in het datacenter landschap?

Als een centrale vorm van computing geen geschikte oplossing blijkt, moet worden bepaald wat dan de beste locatie is voor computing aan de edge. Als we het over edge computing hebben, wordt al snel gezegd dat de edge zo dicht mogelijk bij de verbonden apparaten dient te zijn. Toch kan dat per toepassing verschillen. De latency tolerantie van de toepassing bepaalt hoe ver van het eindpunt de dataverwerking kan plaatsvinden. Dat kan variëren van enkele honderden meters tot enkele honderden kilometers. Vervolgens is het een kwestie van uitrekenen welke oplossing het meeste kosten-efficiënt is. De volgende opsomming biedt inzicht in de belangrijkste mogelijkheden:

- **On-site:** als we het hebben over een boerderij met koeien die met een slimme halsband rondlopen, kan de server op de boerderij staan, als eigendom van de boer, of gehuurd van de aanbieder van de toepassing. We zien ook dat grote cloud providers als AWS en Microsoft oplossingen hiervoor in de markt zetten, waarbij je de cloud in een soort koffer, een draagbaar microdatacenter, kunt meenemen naar een locatie. Dergelijke on-site oplossingen kunnen we ook tegenkomen in fabrieken, olievelden, winkelcentra, treinstations of openbare ruimten.

- **Microdatacenters:** de algemene verwachting is dat er in de komende jaren netwerken van microdatacenters ontstaan met een onderlinge afstand van maximaal een kilometer of 10. Zo hebben bijvoorbeeld mobiele zendmasten meer dan genoeg ruimte voor microdatacenters en kunnen ze zo'n ultra-lokale colocatierol²⁸ invullen.
- **Regionale datacenters:** ook regionale datacenters op een maximale afstand van een kilometer of 50 tot de locatie kunnen hierin een rol spelen, dankzij hun lokale aanwezigheid. Belangrijke voordelen van een regionaal datacenter²⁹ zijn de relatief eenvoudige toegankelijkheid, goede beveiliging en hoge beschikbaarheid.
- **Metrodatacenters**³⁰: naast de micro- en regionale datacenters, rekenen we ook de metrodatacenters tot de edge. Metrodatacenters zijn datacenters waar veelal lokale hubs huizen van grote cloud en content providers. Ze bedienen met deze datacenters een grootstedelijk gebied. In Nederland valt dat niet echt op, want qua omvang is Nederland vergelijkbaar met een enkel grootstedelijk gebied.



Bron: Pb7 Research, 2020

²⁸ Colocatie: het huren van geconditioneerde datavloer en ruimte in computerracks bij een externe partij.

²⁹ Regionaal datacenter: datacenter dat zich voornamelijk bezighoudt met diensten voor klanten binnen een actieradius van circa 30 minuten.

³⁰ Metrodatacenter: datacenter dat als een hub in een grootstedelijke regio is geplaatst om deze regio met een betere performance te kunnen bedienen.

Samenvattend

Edge computing neemt een steeds belangrijkere positie in binnen het snelgroeiende Internet of Things. Het biedt de mogelijkheid om handelingen razendsnel te automatiseren en vormt een oplossing voor 'Things' die veel data creëren, waarvan het grootste deel alleen lokaal nodig is. Voor edge computing zijn veel aansprekende use cases die een belangrijke rol spelen in de digitale transitie van veel organisaties. Steeds meer bedrijven zullen edge computing de komende jaren dan ook omarmen. Daarbij kunnen ze kiezen uit verschillende typen datacenters.

ROL VAN HET REGIONALE DATACENTER



ROL VAN HET REGIONALE DATACENTER

Zoals beschreven, heeft data gravity een groeiende invloed op enterprise computing. Organisaties gaan steeds meer vanuit de behoeften van data denken, hoe ze hun systemen daarop kunnen inrichten en hoe edge computing een deel van die problemen kan oplossen. Intussen is de uitdaging om al deze gedistribueerde landschappen als één virtueel datacenter te laten functioneren. Bevindt het nieuwe knooppunt van dit virtuele datacenter zich wellicht in het regionale datacenter? We proberen die vraag in dit hoofdstuk te beantwoorden door eerst te kijken naar hoe het datacenterlandschap aan het veranderen is. Vervolgens onderzoeken we de mogelijke rol van regionale datacenters in relatie tot de edge. Tot slot beschouwen we hun rol voor enterprise computing als geheel.

Rol van het regionale datacenter

Nederland is een echt datacenterland. De internationale datahub in en rond Amsterdam en de hyperscale datacenters³¹ van Microsoft en Google trekken daarbij meestal de meeste aandacht. Minder bekend, maar net zo bijzonder, is dat Nederland beschikt over een unieke dekking van datacenters op regionaal niveau. Voor meer dan 90% van alle Nederlandse bedrijven zijn er een of meerdere carrier neutrale, hoogkwalitatieve multi-tenant datacenters³² beschikbaar binnen een half uur rijden vanaf kantoor. Dat biedt al deze bedrijven de mogelijkheid om een gedegen 'make-or-buy' beslissing te nemen voor datacenterruimte.

³¹ Hyperscale datacenter: zeer groot datacenter van een cloud provider, van waaruit deze alleen zijn eigen diensten aanbiedt.

³² Multi-tenant datacenter: datacenter waar meerdere huurders gebruik van kunnen maken, bijvoorbeeld door middel van colocation.

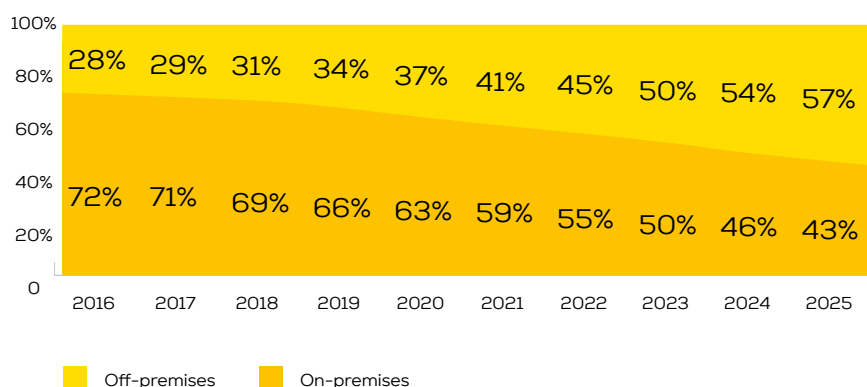
Doordat er steeds hogere eisen worden gesteld aan computerruimten, vooral in het kader van continuïteit in een digitale 24 uurseconomie, is dat geen overbodige luxe. We zien wel dat de rol van datacenters in de regio aan het veranderen is, mede door de effecten van data gravity.

Cloud en het datacenter

De opkomst van de cloud heeft een grootschalige migratie op gang gebracht van applicaties en databases. Dat heeft in het kort de volgende effecten op het datacenterlandschap:

1. De behoefte aan vierkante meters datavloer neemt af bij bedrijven.
2. De business case voor datacentervernieuwing wordt moeilijker bij faciliteiten die aan het eind van hun levensfase zijn. Dat geeft een sterke impuls aan de colocationmarkt.
3. Het verplaatsen van workloads van eigen systemen naar de cloud betekent dat de colocation footprint afneemt.
4. Datacenters worden steeds meer gevuld door aanbieders van SaaS, publieke IaaS/PaaS, managed cloud, content en digitale diensten, die immers steeds meer ruimte nodig hebben.

Verhouding totaal aantal workloads off-premises (colocatie, cloud, hosting) versus on-premises, Nederland 2016-2025



Bron: Pb7 Research, 2020

Vijf jaar geleden vertegenwoordigden bedrijfsdatacenters³³ (single tenant³⁴) meer dan de helft van alle vierkante meters datavloer. In 2020, zo blijkt uit lopend onderzoek van Pb7 Research, beschikken multi-tenant datacenters en hyperscalers gezamenlijk over tweemaal zoveel oppervlakte als single-tenant datacenters. De reden achter deze snelle verandering in verhouding is dat multi-tenant datacenters en vooral de hyperscale datacenters zeer sterk zijn gegroeid, ten opzichte van een lichte daling van bedrijfsdatacenters.

³³ Bedrijfsdatacenter: single tenant datacenter dat (vrijwel) alleen voor eigen doeleinden wordt gebruikt; anders dan een hyperscale datacenter.

³⁴ Single tenant datacenter: datacenter die alleen door de uitbater zelf gebruikt wordt. Dat kan zowel een bedrijfsdatacenter als een hyperscale datacenter zijn.

Het datacenter als knooppunt

Hoewel de groei voornamelijk plaatsvindt in de publieke cloud, kiezen de meeste organisaties er bewust voor om ook een on-premises footprint aan te houden. Het kan zijn dat het niet kosteneffectief is om bepaalde workloads naar de cloud te verplaatsen of dat men bepaalde data liever binnenshuis houdt. Denk daarbij aan privacygevoelige data van patiënten, of concurrentiegevoelige data zoals chemische formules. De meeste van deze bedrijven gaan vervolgens op zoek naar een goede hybride (cloud) oplossing.

Voor regionale datacenters betekent dit dat er steeds meer van ze wordt verwacht. Ze fungeren niet langer alleen als het nabijgelegen alternatief voor een eigen datacenter. Het regionale datacenter wordt voor zakelijke klanten het knooppunt tussen verschillende datalocaties. In het datacenter worden de workloads die on-premises blijven, gekoppeld aan publieke clouds via een beveiligde, snelle toegangspoort. Nu bedrijven steeds meer edge computing omarmen, moeten ook deze datastromen in het regionale datacenter samenkomen. Door gebruik te maken van een netwerk van regionale datacenters, kunnen edge workloads worden verzorgd en landelijk aan elkaar worden gekoppeld.

Datacenters en de edge

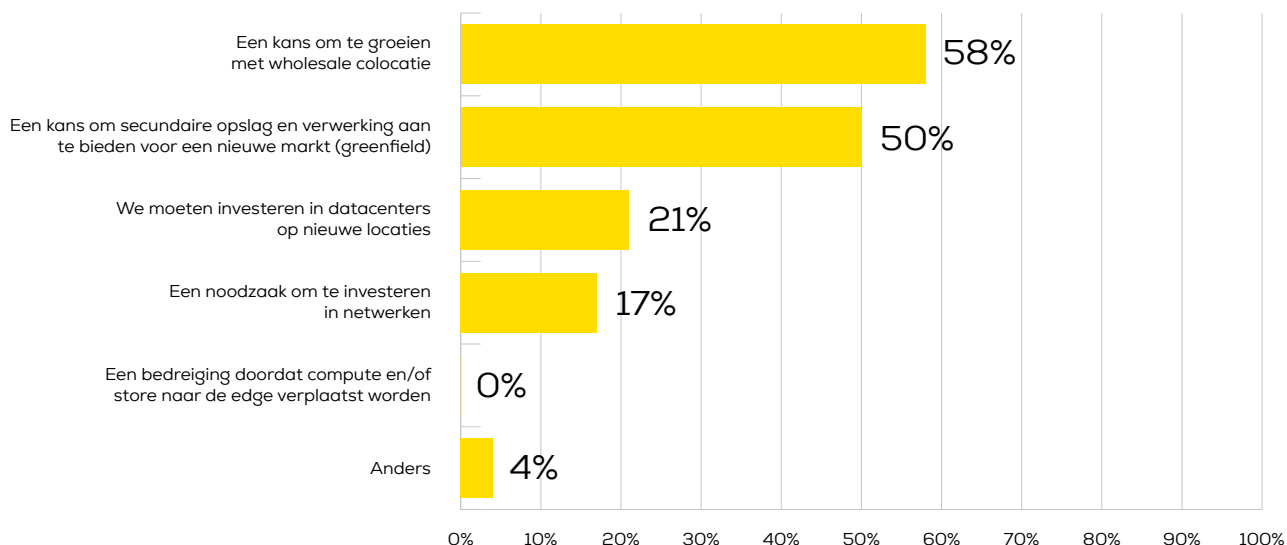
Voor datacenters vormt de opkomst van edge computing voorlopig zeker geen bedreiging, zoals een uitvraag onder de leden van de Dutch Data Center Association laat zien (Pb7 Research, 2019). Sterker nog, veel datacenters zien mogelijkheden in een rol als metro-datacenter en richten hun pijlen daarom sterker op wholesale colocatie³⁵.

Datacenters in de regio spelen in dit grotere wholesale marktsegment niet direct een rol. Ze zien, net als andere colocatie-aanbieders, wel andere mogelijkheden om hiervan te profiteren. Ze gaan er terecht vanuit dat de meeste edge computing workloads, nieuwe workloads zijn. Naast het gegeven dat edge computing veelal heel dicht bij het eindpunt plaatsvindt, worden er ook veel data naar meer centrale datacenters toegestuurd. Daarbij gaat het om data voor rapportagedoeleinden of bijvoorbeeld een bewaarplicht of -wens. Een datacenter in de regio is in sommige gevallen bovendien een prima locatie voor de node/gateway en router waar de data van de apparaten naar toe worden gestuurd.

³⁵ Wholesale colocatie: colocatie gericht op klanten met een vermogen van 1 MW of meer.

Wat betekent de opkomst van 'edge computing' voor datacenteraanbieders als de uwe? Mei 2019

(N=24, leden van de Dutch Data Center Association).



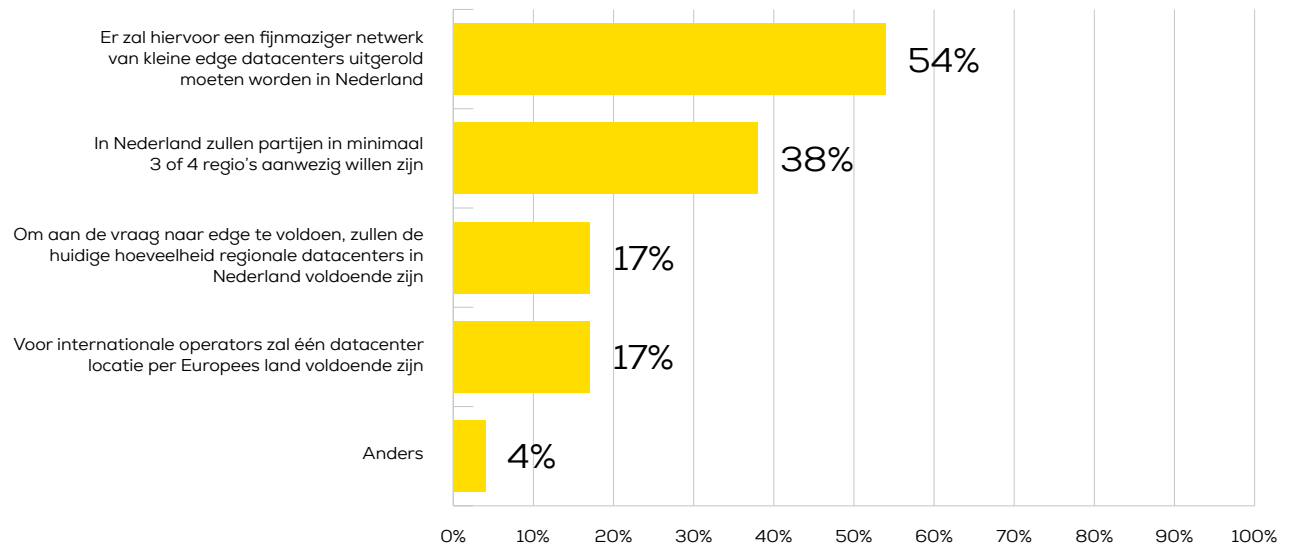
Bron: Pb7 Research, 2019

Veel Nederlandse datacenteraanbieders denken dat het huidige aantal regionale datacenters niet zal voldoen om in te spelen op de opkomende vraag naar edge computing. Regionale datacenters zullen minimaal in drie of vier regio's over een locatie moeten beschikken, willen ze meedoen in deze markt. Er is, zo verwacht men, uiteindelijk een nog fijnmaziger netwerk nodig, bestaande uit een groot aantal ultra-lokale microdatacenters.

Maar het lijkt niet waarschijnlijk dat de huidige generatie datacenters daarop in gaat springen. Uitbaters van bijvoorbeeld antennes voor mobiel spraak- en dataverkeer zijn een logische kandidaat voor ultra-lokale microdatacenters, maar er staan ook starters klaar om in dit gat te duiken. Het lijkt in Nederland echter nog geen uitgemaakte zaak dat er ook echt veel vraag naar ultra-lokale multi-tenant oplossingen gaat ontstaan. Een combinatie van ultra-lokale cloudoplossingen, of all-in-one leveranciersoplossingen, in combinatie met regionale datacenters ligt voorlopig meer voor de hand.

Hoe ziet u de impact van 'edge computing' op het datacenter netwerk? Mei 2019

(N=24, leden van de Dutch Data Center Association).



Bron: Pb7 Research, 2019

Samenvattend

Door data gravity wordt het zwaartepunt van een IT-omgeving verplaatst van binnen naar buiten de organisatie. Bovendien is er niet langer sprake van een enkel zwaartepunt, maar worden dat er steeds meer. Het is belangrijk om te erkennen dat er een fysiek knooppunt nodig is die aan hoge eisen voldoet. Met de afnemende rol van het eigen bedrijfsdatacenter, ligt een extern datacenter voor de hand. Datacenters in de regio ontwikkelen zich steeds meer tot het bedrijfsdatacenter van bedrijven en instellingen. Ze koppelen bedrijfsomgevingen veilig en eenvoudig aan publieke clouds en snelle netwerken en werken zij aan zij met cloudaanbieders. Wat edge computing betreft, zijn de kaarten nog niet geschud. Veel nodes/gateways bevinden zich op het terrein van de gebruiker. Mogelijk zien we een of meerdere netwerken van ultra-lokale microdatacenters ontstaan of blijken de bestaande regionale datacenters juist de meest effectieve multi-tenant oplossing te kunnen bieden.

ANALYSE EN AANBEVELINGEN



ANALYSE

Data gravity is een leidende factor aan het worden in het IT-landschap van Nederlandse bedrijven en instellingen. Maar de vraag is of we de hypothese kunnen bevestigen die we aan het begin van dit onderzoek hebben geformuleerd: ‘door data gravity is gedistribueerde computing weer helemaal terug’? Om die hypothese te toetsen, hebben we een vijftal onderzoeksvragen geformuleerd en proberen te beantwoorden.

1. In welke mate groeit de databerg?

De databerg groeit al jaren exponentieel. Ook voor de komende jaren verwacht IDC dat de databerg elke drie jaar een verdubbeling zal laten. De belangrijkste redenen zijn het toenemend gebruik van videotoeepassingen en streaming, nieuwe onlinediensten en de explosie van het Internet of Things. In 2025 hebben we het inmiddels over 175 zetabytes (10²¹). Een aanzienlijk deel van die data zitten al in de cloud. Dat deel zal de komende jaren verder toenemen ten koste van de data die we nu nog in bedrijfsdatacenters aantreffen.

2. In welke mate worden toepassingen naar de data toe getrokken?

Het verplaatsen van een kleine hoeveelheid data kan razendsnel. Maar het verplaatsen van grote hoeveelheden data kost veel tijd en geld. Hoe meer data er worden opgeslagen en gecreëerd in de cloud, hoe meer de applicaties en services die de data wensen te gebruiken ook naar de cloud gaan. Het is niet altijd logisch en efficiënt om data centraal op te slaan of te verwerken. Zeker niet als op een ver weg gelegen locatie een apparaat op een snelle respons wacht. Dan werkt data gravity ineens omgekeerd. Ver van de cloud is de zwaartekracht van het apparaat het sterkst en worden applicaties en services op of bij devices geplaatst.

3. Waar bevinden zich onze data, hoe verandert dat onder invloed van data gravity en wat is de impact daarvan op enterprise computing en de cloud?

Vanuit het perspectief van een Nederlandse organisatie bevindt het grootste deel van de data, 63% in 2020, zich nog altijd binnen bedrijfseigen serverruimtes en datacenters. Dat duurt niet lang meer. De vloeroppervlakte van de in Nederland aanwezige multi-tenant datacenters plus hyperscale datacenters is al twee keer zo groot als de on-premises datavloer. Deze verzorgen niet alleen de workloads van en voor bedrijven in Nederland. Hyperscale datacenters en grote internationale datacenters in en rond Amsterdam verzorgen ook een groot aantal internationale workloads. Doordat steeds meer data in de cloud zitten, lopen bedrijfsdatacenters leeg en kiezen organisaties vaker voor colocatiedatacenters. Daar bouwen ze aan een hybride (cloud) omgeving. Het datacenter wordt daarmee de digitale toegangspoort tot de wereld. Als we het vervolgens over de edge hebben, zien we dat (nog altijd) relatief grote hoeveelheden data in de buurt van devices ontstaan, daar worden verwerkt en (soms) ook daar worden opgeslagen.

4. Wat betekent data gravity voor de IT-architectuur en hoe houd ik dat beheersbaar?

Organisaties erkennen impliciet of expliciet de noodzaak om in de IT-architectuur rekening te houden met data gravity. Ze beginnen over te stappen van een applicatie-centrische visie naar een data-centrische strategie. De noodzaak komt vooral voort uit de silovorming die, paradoxaal genoeg, is ontstaan door de (workload voor workload) gang naar de cloud. Daardoor dreigen organisaties hun grip kwijt te raken op de eigen data en komt de veiligheid en continuïteit in gevaar. Een betere coördinatie en bijbehorende governance structuur dienen prioriteiten te zijn. Maar ook zal er ingegrepen moeten worden in de organisatie, door een Chief Digital (of Data) Officer aan te wijzen die een mandaat heeft om regels rond data af te dwingen.

5. Wat is de rol van nationale datacenters met een regionale dekking?

Het regionale datacenter vormt steeds meer het knooppunt in een decentrale IT-architectuur. De data uit eigen systemen ontmoeten hier op een veilige en betrouwbare wijze de cloud- en hostingdiensten die worden afgenomen. De groeiende informatiestroom van apparaten uit de edge voegt zich daar steeds meer bij. Hoe groot de rol van regionale datacenters precies zal zijn bij de dataverwerking en aansturing van die apparaten, is nog niet uitgekristalliseerd. Nationale datacenters met een regionale dekking zijn daar het beste voor gepositioneerd.

Is door data gravity gedistribueerde computing weer helemaal terug'?

Wat betekenen deze uitkomsten voor de hypothese? In eerste instantie zien we dat data gravity ook in de komende jaren tot een verdere centralisatie van data leidt. Maar we zien tevens dat de cloud weer steeds dichterbij ons toe komt. Grote cloud- en online contentaanbieders vinden we inmiddels in Nederland terug, met eigen hyperscale datacenters of in gehuurde ruimtes bij multi-tenant datacenters. En juist deze grote cloudaanbieders zetten agressief in op edge-oplossingen, een soort cloud in een koffer oplossing, waarbij de verbinding met de grote publieke cloud gewoon uit kan staan. We krijgen te maken met een computing realiteit waar we centraal doen wat we centraal kunnen doen, lokaal doen wat we lokaal moeten doen en goed nadenken over wat zinvolle, tussenliggende datahubs zijn. Met de explosieve groei van miljarden apparaten en toepassingen aan de edge, kunnen we de hypothese volmondig bevestigen.

AANBEVELINGEN

We sluiten dit rapport af met een aantal aanbevelingen. De eerste set aanbevelingen richt zich op de invloed van data gravity op het gebied van de veranderende wereld van enterprise computing. Daarna volgt een set aanbevelingen voor organisaties die stappen willen maken op het gebied van edge computing.

Enterprise computing

- **Stel de data centraal.** Erken niet alleen dat data heel belangrijk zijn voor de organisatie, maar handel daar ook naar. Voor de cloudstrategie betekent dit bijvoorbeeld dat applicaties en services in de buurt van de data worden geplaatst. En dat de data indien nodig dicht bij de gebruiker worden geplaatst.
- **Pas de organisatie erop aan.** Maak duidelijk dat niet de gebruiker of de budgethouder voor een applicatie eigenaar van de data is, maar de organisatie als geheel. Organisatorisch kan je dat alleen maar doorbreken door een Chief Digital Officer (CDO) aan te stellen. Dan moet deze CDO wel het mandaat hebben om ook daadwerkelijk te fungeren als een Chief Digital Officer. Dat mandaat moet zich bovendien uitstrekken tot alle databronnen, dus ook die uit Operationele Technologie (OT³⁶) en IoT. Houd er rekening mee dat dat niet zonder slag of stoot zal gaan, want IT en OT zijn twee heel verschillende werelden. OT heeft zijn eigen werkwijzen en levenscycli en er moet een inhaalslag worden gemaakt op het gebied van security. Zorg ook voor goede afspraken met betrekking tot data governance.
- **Verminder de complexiteit.** Veel organisaties hebben een lange periode achter de rug waarbij relatief veel ruimte was voor experimenteren en clouddiensten konden worden aangeschaft buiten de IT-afdeling om. De wildgroei aan cloudtoepassingen die zo is ontstaan, moet aan banden worden gelegd. Dat gebeurt door heldere keuzes te maken omtrent IaaS/PaaS platforms en door heldere richtlijnen op te stellen voor de selectie van SaaS-leveranciers en eisen op het gebied van bijvoorbeeld interoperabiliteit waar SaaS aan dient te voldoen. Tenslotte, richt een governance organisatie in die de applicatiebouwers en beheerders helpt om volgens de richtlijnen te werken in de vorm van 'good practises' en 'best practices'.

³⁶ OT: staat voor Operationele Technologie. Technologie om machines (zoals industriële omgevingen) aan te sturen.

- **Blijf beweeglijk.** Als data moeilijk van hun plek te krijgen zijn, zorg er dan voor dat applicaties en services dat niet zijn. Investeer in kennis over microservices, containers en container orkestratie en voorkom lange implementatietrajecten waar applicaties ‘in beton worden gegoten’.
- **Word een knooppunt.** De IT-afdeling heeft twee belangrijke functies: 1. de IT-infrastructuur inrichten en 2. de IT-infrastructuur in de lucht houden. Door het toenemende gebruik van cloud en hosting hoeft de IT-afdeling steeds minder zelf te bouwen, maar moet het wel ketens over leveranciers heen in de gaten gaan houden. Focus op het maken van eenvoudig te gebruiken verbindingen tussen verschillende bronnen, tussen onder meer de public cloud en on-premises systemen en investeer in passende beheertools.
- **Stel de juiste eisen aan het datacenter.** Ook het colocatiedatacenter waar een organisatie gebruik van maakt, valt onder de ketenverantwoordelijkheid. Kies voor een datacenter dat carrier- en cloudneutraal is, waarmee je op een snelle en veilige manier kan koppelen aan netwerken van verschillende cloudaanbieders. Kies voor een datacenter met een landelijke dekking als edge workloads over het land zijn verspreid.

Edge computing

- **Experimenteer met mate.** Om optimaal te profiteren van nieuwe technologie is het altijd belangrijk dat er ruimte voor experimenten is. Maar een experiment met te weinig eisen aan belangrijke zaken als veiligheid, interoperabiliteit en continuïteit, zal nooit op een goede manier kunnen opschalen. Start dus vroeg met het opstellen van een set duidelijke richtlijnen en harde minimale eisen, maar behoud de capaciteit om de richtlijnen aan te passen aan de lessen uit de praktijk.
- **Pas ‘Secure by design’ toe.** Zorg ervoor dat veiligheid en security in de edge en (andere) IoT-oplossingen vanaf het allereerste begin zijn meegenomen. Veiligheid achteraf proberen toe te voegen is ook hier vragen om problemen. Zorg ervoor dat de ontvangende apparaten zich goed kunnen authenticeren voordat ze data kunnen ontvangen. Gebruik verder kunstmatige intelligentie (AI) om te bepalen of er niet gerommeld is met de data die ontvangen worden, zodat onveilige en onlogische commando's niet worden opgevolgd. Tenslotte, zorg voor adequate netwerkbeveiliging, zodat data veilig over de diverse netwerken kunnen reizen, bij voorkeur versleuteld.

- **Kies zorgvuldig de beste passende locatie.** Breng de latency behoeften in kaart en bepaal hoe ver de edge node van het eindpunt af mag staan. Op basis daarvan wordt al snel duidelijk welke huisvestigingsmogelijkheden geschikt zijn, van microdatacenter op locatie tot metrodatacenter. Werk de business case uit van de verschillende mogelijkheden en kies op basis daarvan de beste oplossing. Soms zal het mogelijk zijn om behuizingen te delen met meerdere IoT-oplossingen.
- **Zorg voor een robuuste behuizing.** In een aantal gevallen kunnen edge nodes in een geconditioneerd datacenter worden geplaatst. Maar als het om ultra-lokale locaties gaat, zullen de omstandigheden meer van de hardware eisen. Houd er ook rekening mee dat de apparatuur slecht bereikbaar of zelfs geheel onbereikbaar is voor onderhoud en dat ook onderhoud op afstand niet altijd mogelijk is. Tenslotte moet de apparatuur ook zelfstandig kunnen functioneren als er geen netwerkverbinding is.
- **Houd rekening met secundaire opslag en verwerking.** Bij veel edge oplossingen zal ook op een centrale locatie opslag en verwerking plaatsvinden. In de edge worden alleen de meest noodzakelijke bewerkingen gedaan die om een directe respons vragen. Er is meestal geen ruimte om grote hoeveelheden data langdurig op te slaan. Bovendien kan het zijn dat de data ook centraal aanvullende bewerkingen ondergaan.
- **Investeer in data beheertools.** Als een edge netwerk uit een groot aantal meetpunten bestaat, raakt men het overzicht al snel kwijt. Investeer in beheertools om het overzicht te houden en gericht incidenten te kunnen oplossen.

Ten slotte

Data gravity is in veel opzichten vergelijkbaar met de zwaartekracht tussen voorwerpen of hemellichamen. Het is zo gewoon, dat we er zelden bij stilstaan. Data trekken meer data, applicaties en services naar zich toe. Hoe groter de datamassa wordt, hoe moeilijker het wordt voor data, applicaties en services aan de zwaartekracht te ontsnappen. Maar net als in ons zonnestelsel draait niet alles om de zon: satellieten zitten gevangen in de zwaartekracht van planeten. En dat is wel zo handig.

De aanwezigheid van data gravity heeft een grote impact op de keuzes die we maken in de IT-architectuur en hoe we IT-systemen bouwen. Hoe beter we de werking begrijpen, hoe beter de keuzes die we maken. Organisaties dienen allereerst te bepalen welke locatie voor welke data de meeste waarde hebben. Dat hangt sterk af van waar bewerking van de data nodig en/of het meest zinvol is. Is dat in een cloud, in een centraal datacenter, of op of bij het apparaat. De volgende stap is bepalen welke applicaties en services bij die data nodig zijn. En tenslotte moeten de locaties en datastromen aan elkaar geknoopt worden op een snelle, veilige en betrouwbare manier.

APPENDIX



METHODOLOGIE

Dit rapport baseert zich op bestaande bronnen. De gebruikte publieke bronnen zijn terug te vinden in de appendix van het rapport. Naast publieke bronnen is gebruik gemaakt van de onderzoeksdatabases en marktinzichten van Pb7 Research. Op basis daarvan is onder meer de marktomvang en groeiverwachting van edge computing tot stand gekomen. Pb7 Research is een onafhankelijk Nederlands onderzoeksbureau dat zich specialiseert in de thema's datacenter, IT-security, cloud, IoT en digitale transformatie en is een onderzoekspartner van de Dutch Data Center Association.

DEFINITIES

Bedrijfsdatacenter

Single tenant datacenter dat (vrijwel) alleen voor eigen doeleinden wordt gebruikt; anders dan een hyperscale datacenter.

Cloud

Via internet op aanvraag toegang tot een gedeelde verzameling eenvoudig aanpasbare computermiddelen zoals servers, software en dataopslag. Diensten worden naar behoefte afgenomen zonder hulp van de aanbieder (selfservice), betalen is op basis van werkelijk gebruik.

Colocatie

Het huren van geconditioneerde datavloer en ruimte in computerracks bij een externe partij.

Content Delivery Network (CDN)

Een netwerk van proxy servers die geografisch verspreid zijn over het internet in verschillende datacenters, zodat gebruikers snel en zonder vertraging content kunnen binnenhalen.

Data gravity

Het naar de datalocatie toetrekken van applicaties en services.

Datahub

Als een data lake, maar dan met deels of geheel geharmoniseerde data, indexering en standaard analytische tools.

Data lake

Verzamelpaats van data (kopie) uit een groot aantal bronnen waarbij de data ongestructureerd worden opgeslagen; typisch voor data science mogelijkheden.

Datacenter

Een geconditioneerde ruimte waarin servers en bijbehorende apparatuur worden geplaatst. Ook serverruimtes binnen gebouwen met andere functies rekenen we in dit document tot datacenters.

Datasilo

Een term om aan te geven dat veel data in aparte databases zitten opgesloten.

Edge

De rand van het netwerk; wat niet tot de kern behoort.

Edge computing

Verwerking van data aan de rand van het netwerk, dichtbij (maar niet op) het apparaat.

Edge gravity

De aantrekkende kracht op applicaties en services van data die ver van de core wordt gecreëerd en gebruikt.

DEFINITIES

Edge node/gateway

Edge nodes (ook wel edge gateways genoemd): computers die in de edge data deels of geheel verwerken voor endpoints (apparaten) en bovendien vaak als een cluster kunnen fungeren.

Endpoint (of on-device) computing

Of wel eindpunt, verwerking van data op het apparaat zelf.

Gedistribueerd computing

Bij gedistribueerde computing wordt een enkele taak over meerdere systemen opgesplitst in meerdere subtaken die relatief onafhankelijk berekend en achteraf weer samengevoegd kunnen worden tot een enkel resultaat.

Gefedereerde database

Virtuele database van waaruit query's kunnen worden gedraaid die direct worden opgehaald uit live databases.

Hybride cloud

Cloudomgeving waar een private cloud workload kan worden geschaald naar een publieke cloud, eventueel vice versa.

Hybride IT

Workloads op eigen client/server omgevingen (kunnen) worden geschaald naar een publieke cloud.

Hyperscale datacenter

Zeer groot datacenter van een cloud provider, van waaruit deze alleen zijn eigen diensten aanbiedt.

IaaS

Clouddienst waarbij IT-infrastructuur, zoals opslag en verwerkingscapaciteit als dienst wordt afgenomen.

Internet of Things

Een voorgestelde ontwikkeling van het internet, waarbij alledaagse voorwerpen zijn verbonden met het netwerk en gegevens kunnen uitwisselen.

Kunstmatige intelligentie of Artificial Intelligence (AI)

Het vermogen van een systeem om externe gegevens correct te interpreteren, om te leren van deze gegevens, en om deze lessen te gebruiken om specifieke doelen en taken te verwezenlijken via flexibele aanpassing.

Latency

De vertraging in de dataoverdracht over een datacommunicatienetwerk. Lage (of low) latency is een minimale vertraging.

Machine Learning (ML)

Het onderdeel binnen kunstmatige intelligentie dat zich bezighoudt met de ontwikkeling van algoritmes en technieken waarmee computers kunnen leren.

DEFINITIES

Managed cloud

Dienst waarbij een service provider afzonderlijke publieke en/of private cloudmiddelen als een geheel inricht en beheert.

Metrodatacenter

Datacenter dat als een hub in een grootstedelijke regio is geplaatst om deze regio met een betere performance te kunnen bedienen.

Microdatacenter

Zeer klein, autonoom functionerend datacenter dat op zeer kleine afstand functioneert van endpoint devices (<10 km).

Multicloud

Er worden meerdere clouds (private en/of public) naast elkaar gebruikt voor verschillende doeleinden/workloads.

Multi-tenant datacenter

Datacenter waar meerdere huurders gebruik van kunnen maken, bijvoorbeeld door middel van colocation.

Off-premises

Op een locatie van een dienstverlener (colocation, cloud, hosting, etc.).

On-premises

Op een locatie van de gebruiker (bedrijf of instelling).

OT

Staat voor Operationele Technologie. Technologie om machines (zoals industriële omgevingen) aan te sturen.

PaaS

Clouddienst waarbij een platform als dienst wordt afgenomen en waar applicaties op draaien en op maat ontwikkeld kunnen worden.

Private cloud

Clouddiensten waarbij middelen door verschillende interne klanten (afdelingen, dochters en/of applicaties) worden gedeeld.

Private edge

Edge oplossingen die gehuisd worden op een bedrijfseigen locatie.

Public cloud

Clouddiensten waarbij de middelen door meerdere afnemers (uit verschillende organisaties) worden gedeeld.

Regionaal datacenter

Datacenter dat zich voornamelijk bezighoudt met diensten voor klanten binnen een actieradius van circa 30 minuten.

Retail colocation

Colocation gericht op klanten met een vermogen van minder dan 1 MW.

DEFINITIES

SaaS

Clouddienst waarbij een applicatie als dienst wordt afgenomen.

Silo; Silo gravity

Een term om aan te geven dat silo's (data die in aparte databases zitten opgesloten), als gevolg van data gravity, lastig af te breken zijn.

Single tenant datacenter (ook wel bedrijfsdatacenter)

Datacenter die alleen door de uitbater zelf gebruikt wordt. Dat kan zowel een bedrijfsdatacenter als een hyperscale datacenter zijn.

BRONNEN

IDC (2018)

in opdracht van Seagate, Data Age 2025

IDG (2019)

The Digitization of the World. From Edge to Core
in opdracht van Capgemini, Dutch Cloud Readiness Report 2019

McCrary, Dave (2010)

Data Gravity – in the Clouds (blogpost)

Pb7 Research (2019)

Dutch Data Center Report 2019

State of the Dutch Data Center

Pb7 Research (2020)

diverse databases