



HET DEFINIËREN VAN VIER EDGE-ARCHETYPEN EN HUN TECHNOLOGISCHE EISEN

Introduction

In de afgelopen paar jaar is 'edge computing' een van de meest besproken trends in IT geworden, en terecht. Grand Valley Research verwacht tussen 2018 en 2025 een [samengesteld jaarlijks groeipercentage \(CAGR\) van 41 procent voor edge computing](#). Bijna elke branche onderkent de beperkingen van het ondersteunen van gebruikers via gecentraliseerde IT-infrastructuren. Edge computing brengt opslag en computing dicht bij gebruikers en apparaten.

Omdat de connectiviteit van apparaten en de enorme gegevensvolumes die gebruikers genereren en verbruiken toenemen, is verandering noodzakelijk. Volgens de [Cisco Visual Networking Index](#) bedroeg de omvang van het wereldwijde IP-verkeer in 2016 1,2 zetabyte. In 2021 is dat vrijwel verdrievoudigd, tot 3,3 zetabyte. Daarnaast verwacht Cisco dat het aantal met IP-netwerken verbonden apparaten in 2021 drie keer zo groot als de wereldbevolking is. Dat komt neer op ruim 23 miljard verbonden apparaten in slechts drie jaar. [Andere bedrijven doen soortgelijke voorspellingen](#). Zo verwacht Gartner dat er in 2020 20,8 miljard aangesloten apparaten zijn, IDC 28,1 miljard en IHS Markit 30,7 miljard.

Een groot deel van deze IoT-gegevens zullen gegevens van mobiele sensoren zijn die via draadloze of mobiele netwerken in plaats van via bedrade internetverbindingen worden verzonden. En dat zet de infrastructuur van mobiele netwerken onder druk. Er wordt geschat dat de omvang van het mobiele IP-verkeer in 2021 [zeven keer zo groot zal zijn](#), een groei die dubbel zo hoog is als die van het vaste IP-verkeer.

De veranderingen in de reken- en opslaginfrastructuur die nodig zijn om de slimme en verbonden toekomst te ondersteunen, met name op lokaal niveau, zijn aanzienlijk.

Helaas zijn er weinig of geen bronnen die een uitgebreid overzicht bieden van het edge-ecosysteem. Een nauwkeurige marktanalyse onthult een grote verscheidenheid aan actuele en opkomende gebruikssituaties. Dit komt met name door de ruime definitie van edge computing. Hoewel er veel overeenkomsten zijn tussen de definities, verschillen ze op een aantal belangrijke punten.

Om deze verschillen en hun implicaties voor de ondersteunende infrastructuur beter te begrijpen, analyseerde Vertiv de gebruikssituaties binnen het edge-ecosysteem. Als gevolg van deze analyse hebben we vier hoofd-archetypen geïdentificeerd voor edge-toepassingen:

- Gegevensintensief
- Gevoelig voor menselijke latentie
- Gevoelig voor latentie tussen machines
- Levensbelang

Deze paper biedt een beschrijving van elk archetype met voorbeelden van de gebruikssituaties met de meeste invloed. Daarnaast geven we een overzicht van de connectiviteitseisen voor lokale, stedelijke en regionale hubs, die de edge-transmissielag en kern voorstellen en waarbij soms onderscheid wordt gemaakt tussen edge, fog en cloud computing.

Edge-gebruikssituaties begrijpen

Om de vier archetypen te identificeren, was het eerst noodzakelijk om de gebruikssituaties voor edge-technologie te begrijpen. Het onderzoeksteam van Vertiv heeft meer dan honderd gebruikssituaties voor edge-technologie geïdentificeerd en beoordeeld. Uiteindelijk is deze lijst voor een gedetailleerdere analyse teruggebracht tot de 24 situaties die de grootste invloed hebben op IT-infrastructuur.

Deze analyse keek naar de prestatie-eisen van elke gebruikssituatie op het gebied van latentie, beschikbaarheid en verwachte groei. Maar ook naar beveiligingseisen, zoals de noodzaak voor encryptie, verificatie en het voldoen aan wettelijke eisen. Verder evalueerden de Vertiv-experts de behoefte aan integratie met bestaande of legacy-toepassingen en andere gegevensbronnen, en het aantal potentiële benodigde locaties om de gebruikssituatie te ondersteunen.

Ook heeft het team de gegevenskenmerken van de gebruikssituaties bestudeerd. Uit die resultaten blijkt dat de toepassingen die elke situatie ondersteunen, in aanvulling op de eisen voor beschikbaarheid en beveiliging, ook eisen voor de werkbelasting hebben waarin gegevens centraal staan. Deze omvatten het gegevensvolume, hoe gegevens worden benaderd, eisen voor gegevensoverdracht, gegevensintegriteit en gegevensanalyse. Deze op gegevens gerichte benadering is gefilterd met eisen voor beschikbaarheid en beveiliging en staat centraal om de eisen van verschillende gebruikssituaties te begrijpen en te categoriseren.

Een lijst van de 24 gebruikssituaties, geordend op archetype, is te vinden in afbeelding 1.

Het edge-ecosysteem

GEGEVENSINTENSIEF	GEVOELIG VOOR MENSELIJKE LATENTIE	GEVOELIG VOOR LATENTIE TUSSEN MACHINES	LEVENSBELANG
<ul style="list-style-type: none"> • Beperkte Connectiviteit • Slimme Steden • Slimme Fabrieken • Slimme Woning/Gebouw • Distributie HD Content • High Performance Computing • Virtual Reality • Olie en Gas Digitalisering 	<ul style="list-style-type: none"> • Slimme Beveiliging • Smart Grid • Lage Latentie Content-Dist • Arbitrage Markt • Analyses in Real Time • Militaire oorlogssimulaties 	<ul style="list-style-type: none"> • Digitale Gezondheid • Verbonden / Autonome Voertuigen • Drones • Slim Vervoer • Autonome Robots 	<ul style="list-style-type: none"> • Website Optimalisatie • Augmented Reality • Slimme Retail • Verwerking Gesproken Taal

Afbeelding 1: Archetypen

Archetype 1: Gegevensintensief

Bandbreedte	Latentie	Beschikbaarheid	Beveiliging
Hoog	Gemiddeld	Hoog	Gemiddeld

Het archetype Gegevensintensief representeert gebruikssituaties waarbij de hoeveelheid gegevens het onpraktisch maakt om gegevens direct via het netwerk naar de cloud of van de cloud naar het gebruikspunt te versturen. Dit kan zijn vanwege het gegevensvolume, de kosten of de bandbreedte.

De distributie van hd-content is waarschijnlijk het meest bekende voorbeeld van een gegevensintensieve edge-toepassing. [In 2016 was video verantwoordelijk voor 73 procent van alle IP-verkeer. De verwachting is dat dit in 2021 zal toenemen tot 82 procent](#), omdat het streamen van video en virtual reality blijven groeien. Belangrijke contentproviders zoals Amazon en Netflix werken actief samen met co-locatieaanbieders om hun netwerken te kunnen uitbreiden en om gegevensintensieve videostreaming dichterbij gebruikers te brengen om zo kosten en latentie te beperken.

Op dit moment wordt reeds [35 procent van de content die wordt geopend door internetgebruikers in Noord-Amerika verzonden vanuit de gemeente waar de gebruiker zich bevindt](#). Naar verwachting zal dit aantal stijgen tot 51 procent in 2021 naarmate aanbieders van content hun netwerken naar de edge blijven uitbreiden. Toch is dit nog maar de eerste golf van core-to-edge computing. Naarmate de vraag naar hd-video groeit, zullen lokale hubs steeds meer de huidige stedelijke hubs ondersteunen om de bandbreedtekosten en latentieproblemen verder te beperken.

Een ander goed voorbeeld van dit archetype is het gebruik van IoT-netwerken om slimme woningen, gebouwen, fabrieken en steden te creëren. In 2018 bleek uit een onderzoek door 451 Research en Vertiv dat, terwijl slechts 33 procent van de 700 onderzochte organisaties IoT op grote schaal toepaste, 56 procent aangaf dat ten minste een kwart van hun IT-capaciteit op dit moment IoT ondersteunt. Ondanks dat IoT zich op dit moment nog in de beginfase bevindt, hebben organisaties nu al moeite om het gegenereerde gegevensvolume te beheren.

De uitdaging is in dit geval tegenovergesteld aan die van contentdistributie. In plaats van gegevens dichterbij de gebruiker te brengen, moeten deze toepassingen de enorme hoeveelheden gegevens die door apparaten en systemen bij de bron worden gegenereerd voor verwerking naar een

centrale locatie sturen. Dit vraagt om de ontwikkeling van een edge-to-core netwerkarchitectuur.

IoT en het Industrial Internet of Things (IIoT) representeren een raster van sensoren die elk uur enorme hoeveelheden gegevens produceren. Deze gegevens ondersteunen een "sense-infer-react"-lus die zichtbaarheid en regeling mogelijk maakt van huishoudelijke apparatuur thuis tot industriële uitrusting. Alleen een subset van deze gegevens wordt overgebracht naar een lokaal, regionaal of cloud-datacenter voor verdere verwerking. Dit houdt in dat enorme hoeveelheden berekeningen vereist zijn aan het uiteinde van de edge om apparaten en systemen in staat te stellen beslissingen te nemen en te reageren op de gegevens die door sensoren zijn geleverd.

De meest eenvoudige van deze toepassingen, de slimme woning, moet meerdere gegevensintensieve apparaten en systemen ondersteunen, waaronder entertainment, airco-systemen en beveiliging.

Gegevensintensief

Volgens IHS Markit [zal de wereldmarkt voor verbonden thuisapparatuur stijgen van meer dan 100 miljoen eenheden in 2017 tot ongeveer 600 miljoen eenheden in 2021](#).

Slimme steden en fabrieken hebben te maken met gegevensuitdagingen die inherent zijn aan die van slimme woningen, maar dan op grote schaal. Veel steden werken reeds met pilots of evalueren smart city-technologie om verkeersstromen te verbeteren, hulpdiensten te ondersteunen en kosten te beperken.

Slimme fabrieken, die gebruikmaken van de convergentie van IoT, cyber-fysieke systemen en cloud computing om fabrikanten in staat te stellen real-time gegevens te gebruiken om de efficiëntie te verbeteren, kosten te beperken en zich aan te passen aan veranderingen in de vraag, worden gezien als de volgende industriële revolutie. Volgens McKinsey hebben fabrieken en andere productie-omgevingen de mogelijkheid om de grootste financiële impact te realiseren van de toepassing van IoT. Ze voorspellen dat IoT in 2025 een [economische waarde zal genereren tussen de 1,2 en 3,7 biljoen dollar](#). Deze waarde zal afkomstig zijn van nieuwe energiezuinigheid, werkproductiviteit, voorraadoptimalisatie en verbeterde veiligheid op het werk. Echter vereist de realisatie hiervan een robuuste infrastructuur.

In de olie- en gasindustrie heeft digitalisering al een enorme verbetering teweeggebracht in de efficiëntie van

onderzoeks- en extractie-processen, maar het heeft ook gezorgd voor enorme uitdagingen op het gebied van gegevensbeheer. Een enkele boorinstallatie kan dagelijks terabytes aan gegevens genereren.

Andere gebruikssituaties die onder het archetype Gegevensintensief vallen, zijn virtual reality, high-performance computing en omgevingen met beperkte connectiviteit, zoals gebieden waar reddingsoperaties plaatsvinden na een natuurramp of cyberaanval.

Wat al deze gebruikssituaties gemeen hebben, is de noodzakelijkheid om grote hoeveelheden gegevens over te brengen naar gebruikers die ze kunnen verwerken, of van apparaten en systemen waar ze zijn gegenereerd naar een centrale opslag.

Archetype 2: Gevoelig voor menselijke latentie

Bandbreedte	Latentie	Beschikbaarheid	Beveiliging
Gemiddeld	Hoog	Gemiddeld	Gemiddeld

Het archetype Gevoelig voor menselijke latentie omvat gebruikssituaties waarbij diensten zijn geoptimaliseerd voor gebruik door de mens. Zoals de naam al suggereert, is snelheid een belangrijke eis voor dit archetype.

De uitdaging van menselijke latentie ligt in de gebruikssituatie van het optimaliseren van klantervaringen. In toepassingen zoals e-commerce heeft snelheid een rechtstreeks effect op de ervaring van een gebruiker: websites die geoptimaliseerd zijn voor snelheid met lokale infrastructuren zien direct het aantal bezoekers van hun website stijgen, en daarmee hun omzet.

Gevoelig voor menselijke latentie

Google heeft vastgesteld dat een extra vertraging van 500 milliseconden in de reactietijd van een pagina 20 procent minder verkeer opleverde en Yahoo stelde vast dat een vertraging van 400 milliseconden 5 tot 9 procent minder verkeer opleverde.

Dit effect strekt zich ook uit naar betalingsverwerking. Amazon heeft vastgesteld dat een vertraging van 10 milliseconden in de betalingsverwerking een omzetzak van 1 procent als gevolg had. Gecentraliseerde goedkeuring via een wachtwoord duurde gemiddeld 7 seconden. De overgang naar lokale verwerking bracht de tijd terug naar 600 milliseconden, een verbetering van 6400 milliseconden, waarbij elke 100 milliseconden in potentie 1 procent extra omzet oplevert.

Een ander opkomend voorbeeld van een voor menselijke latentie gevoelige toepassing is de verwerking van gesproken taal. In de toekomst is de stem waarschijnlijk de voornaamste vorm van interactie met dagelijkse IT-toepassingen. Alexa en Siri voeren de verwerking van gesproken taal op dit moment nog in de cloud uit. Naarmate het ondersteunde aantal gebruikers, toepassingen en talen toeneemt, zal het nodig zijn om deze mogelijkheden dichterbij de gebruikers te brengen.

Andere gebruikssituaties van menselijke latentie zijn slimme retail, zoals de kassaloze Amazon Go-winkels, en ingrijpende technologieën zoals augmented reality, waarbij kleine vertragingen het verschil kunnen betekenen tussen plezier en misselijkheid. In elk geval hebben vertragingen in de levering van gegevens rechtstreeks invloed op de ervaring van het gebruik van technologie, zoals bij spraakverwerking en augmented reality, of op de omzet en winstgevendheid van een winkel, zoals bij de optimalisatie van websites en slimme retail. Naarmate deze gebruikssituaties toenemen, zal ook de behoefte aan hubs voor lokale gegevensverwerking toenemen.

Archetype 3: Gevoelig voor latentie tussen machines

Bandbreedte	Latentie	Beschikbaarheid	Beveiliging
Gemiddeld	Hoog	Hoog	Hoog

Het archetype Gevoelig voor latentie tussen machines omvat gebruikssituaties waarbij diensten geoptimaliseerd zijn voor gebruik door machines. Omdat machines gegevens veel sneller kunnen verwerken dan mensen, is snelheid ook het bepalende kenmerk voor dit archetype. De gevolgen van langzame levering kunnen in dit geval zelfs nog groter zijn dan bij het archetype Gevoelig voor menselijke latentie.

Zo zijn bijvoorbeeld de systemen die gebruikt worden bij geautomatiseerde financiële transacties, zoals de handel in grondstoffen en aandelen, latentiegevoelig. In deze situaties kunnen prijzen binnen milliseconden veranderen en systemen die niet over de nieuwste gegevens beschikken wanneer nodig, kunnen transacties niet optimaliseren. Zo kunnen potentiële winsten op verliezen uitdraaien.

Gevoelig voor latentie tussen machines

Volgens een onderzoek van de Tabb Group kan een broker [per milliseconde wel vier miljoen aan inkomsten verliezen](#) als zijn elektronische handelsplatform 5 milliseconden achterloopt op dat van de concurrentie.

Smart grid-technologie valt eveneens onder dit archetype. Deze technologie wordt toegepast in het elektrische distributienetwerk om aanbod en vraag zelfstandig in evenwicht te houden en om het gebruik van elektriciteit op een duurzame, betrouwbare en economische wijze te beheren. Dit stelt distributienetwerken in staat om zichzelf te herstellen, kosten te optimaliseren en onderbroken stroombronnen te beheren. Mits de juiste gegevens op het juiste moment beschikbaar zijn.

Andere voorbeelden van toepassingen die gevoelig zijn voor latentie tussen machines zijn slimme beveiligingssystemen die op beeldherkenning vertrouwen, militaire oorlogssimulaties en analyses in real time.

Archetype 4: Levensbelang

Bandbreedte	Latentie	Beschikbaarheid	Beveiliging
Gemiddeld	Hoog	Hoog	Hoog

Het archetype Levensbelang omvat gebruikssituaties die een directe invloed hebben op de gezondheid van veiligheid van de mens. In deze gebruikssituaties zijn snelheid en betrouwbaarheid cruciaal.

Het beste voorbeeld van het archetype Levensbelang zijn waarschijnlijk autonome voertuigen en drones, die grote voordelen opleveren als ze naar behoren functioneren, maar een gevaar kunnen vormen als ze verkeerde beslissingen nemen.

Autonome voertuigen hebben zich sneller ontwikkeld dan werd verwacht, waardoor een aantal autofabrikanten en technologische bedrijven nu al actief voertuigen testen. Bij de meeste van deze voertuigen zit een persoon klaar op de bestuurdersplaats om, wanneer nodig, automatische bediening te omzeilen en risico's te beperken. In de nabije toekomst zullen bestuurderloze voertuigen en transportsystemen de weg op gaan. Als deze systemen niet over de benodigde gegevens kunnen beschikken op het cruciale moment, kunnen de gevolgen rampzalig zijn.

Hetzelfde geldt voor drones. Een toekomst waarbij op een willekeurig moment honderden bezorgdrones over een stad vliegen is niet moeilijk voor te stellen.

Levensbelang

Grote pakketbezorgers en e-commerce bedrijven zoals Amazon en DHL experimenteren nu al met drones voor het bezorgen van pakketten.

Het toegenomen gebruik van technologie in de gezondheidszorg behoort eveneens tot het archetype

Levensbelang. Elektronische gezondheidsdossiers, cybermedicijnen, gepersonaliseerde medicijnen (genoomkartering) en zelfbewakende apparaten geven nieuwe vorm aan de gezondheidszorg en genereren enorme hoeveelheden gegevens.

Andere voorbeelden zijn slim vervoer en autonome robots. De vervoers- en logistiekbranche is op zoek naar oplossingen waarbij gegevens centraal staan, ter verbetering van de veiligheid van bestuurder en passagiers, brandstofefficiëntie en het beheer van hulpmiddelen. Technologie op dit gebied omvat intelligente vervoerssystemen, vlootbeheer en telematica, geleidings- en regelsystemen, passagiersentertainment en handelstoepassingen, reserverings-, tolheffings- en ticketsystemen, en beveiligings- en bewakingssystemen.

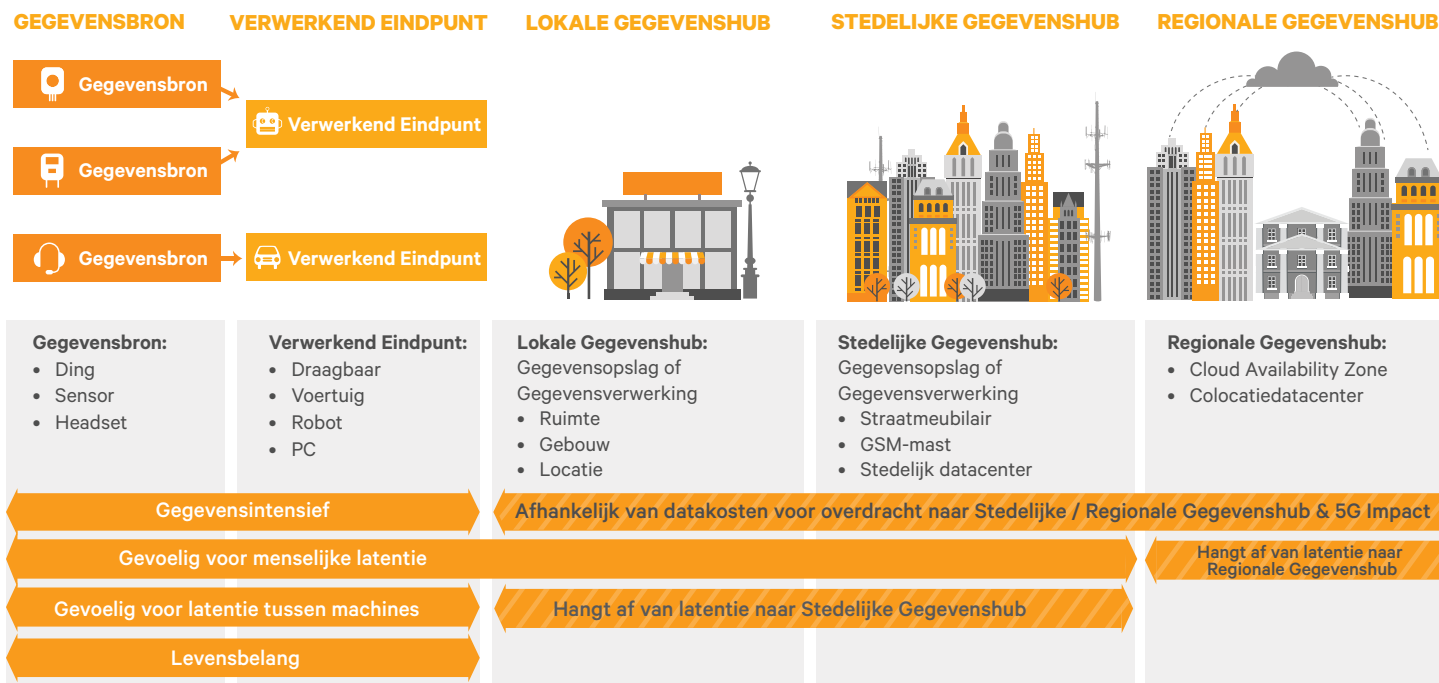
Technologische eisen voor lokale en regionale hubs

De infrastructuur die nodig is om deze actuele en vastgestelde gebruikssituaties te ondersteunen bestaat uit vier lagen opslag en berekening, bovenop de communicatie-infrastructuur die nodig is om gegevens tussen de lagen te verplaatsen.

Aan de bron staat het apparaat dat gegevens genereert of verbruikt en een verwerkend eindpunt. Het apparaat kan een sensor zijn die van alles bewaakt, van de ingeschakelde status van een lamp, de toegang tot een deur, de temperatuur in een kamer of andere gewenste informatie. Het verwerkende eindpunt kan zoiets eenvoudigs zijn als de pc of tablet waar een gebruiker video naartoe stuurt. Tevens kunnen het microprocessors zijn die zijn ingebouwd in auto's, robots of draagbare apparaten. Deze componenten zijn toepassingsafhankelijk en maken gewoonlijk onderdeel uit van het oorspronkelijke ontwerp van een apparaat, of ze worden achteraf toegevoegd aan bestaande apparaten.

Elk archetype, met uitzondering van het archetype Levensbelang, kan afhankelijk van de toepassing in de lokale gegevenshub worden opgenomen. De lokale gegevenshub zorgt voor opslag en verwerking dicht bij de bron. In sommige gevallen kan de lokale hub een zelfstandig datacenter zijn, maar meestal is het een rack- of row-based systeem dat 30 tot 300 kW capaciteit levert in een geïntegreerde behuizing die in elke omgeving kan worden geïnstalleerd.

Deze rack- en row-based behuizingssystemen integreren communicatie, berekeningen en opslag met de bijbehorende stroombeveiliging, klimaatregeling en fysieke beveiliging. Voor archetypen die een hoge mate van beschikbaarheid vereisen, zoals Gevoelig voor latentie tussen machines en



Levensbelang, moet de lokale hub een redundante back-upstroomvoorziening bevatten en voorzien zijn van mogelijkheden voor beheer en bewaking op afstand. Veel gebruikssituaties vereisen tevens gegevensversleuteling en andere beveiligingsmaatregelen binnen de lokale hub.

Voor alle archetypen, afgezien van Levensbelang, kan de stedelijke en/of regionale hub worden gebruikt ter ondersteuning van de gebruikssituaties. Dit is afhankelijk van de kosten van gegevensoverdracht, de bandbreedte die mogelijk is door toepassing van 5G en de latentie naar de locatie van het fysieke datacenter. De stedelijke hub maakt gebruik van de bestaande telefonie-infrastructuur om reken- en infrastructuurcapaciteit te bieden. Deze zal worden ontworpen volgens telefonienormen, inclusief gelijkstroom en koeling met buitenlucht, met ondersteuning voor een veel ruimer temperatuur- en luchtvochtigheidsbereik dan gebruikelijk is in traditionele datacenters. De regionale hub is waarschijnlijk een cloud of colocatiedatacenter dat zich in dezelfde regio als de lokale en stedelijke hub bevindt.

Voor zowel de stedelijke als regionale hubs moeten modulaire ontwerpen worden overwogen die eenvoudig en verder schaalbaar zijn dan de aanvankelijke ontwerpspecificaties, om zo onverwachte toenames in vraag te anticiperen. Deze installaties moeten ook worden ontworpen om schaalbaar te zijn qua dichtheid. Beeldintensieve toepassingen, zoals virtual reality, en verwerkingsintensieve toepassingen, zoals analyse en

machine learning, vereisen waarschijnlijk rackdichtheden die de gebruikelijke 10 kW ontwerpspecificatie overschrijden. In vrijwel alle gevallen moeten deze hubs hetzelfde of een hoger niveau aan redundantie en beveiliging bieden als de lokale hub.

Vooruit kijken

Door de behoeften aan werkbelasting te bepalen voor de 24 besproken gebruikssituaties zijn vier belangrijke archetypen vastgesteld. Deze archetypen kunnen een leidraad vormen voor beslissingen met betrekking tot infrastructuur- en configuratie-eisen, voor de geanalyseerde gebruikssituaties alsmede voor situaties die zich in de komende jaren voor zullen doen. Vertiv bouwt voort op dit initiële onderzoek om de specifieke technologische eisen en configuraties voor elk archetype verder te definiëren.



VertivCo.nl | Vertiv Benelux BV, Druivenstraat 3, 4816 KB Breda, Nederland

© 2018 Vertiv Co. Alle rechten voorbehouden. Vertiv en het Vertiv-logo zijn gedeponeerde handelsmerken van Vertiv Co. Alle andere namen en logo's waarnaar wordt verwezen, zijn handelsnamen, handelsmerken of gedeponeerde handelsmerken van de betreffende eigenaren. Hoewel alle voorzorgsmaatregelen zijn getroffen om de juistheid en volledigheid te waarborgen, aanvaardt Vertiv Co. geen verantwoordelijkheid en wijst alle aansprakelijkheid af voor fouten of omissies, of voor schade die het gevolg is van deze informatie. Specificaties zijn onderhevig aan verandering zonder voorafgaande kennisgeving.