

imec.istart

Tengu: de toolbox voor snelle en eenvoudige Big Data-analyse

Qrama, een imec spin-off die ook opgenomen is in het imec.istart-programma, ontwikkelde het Tengu-platform: een generisch open-source platform dat professionals volledige vrijheid geeft om de beste tools te kiezen voor elk van hun data science-projecten.

In een wereld gedreven door Big Data ervaren datawetenschappers een grote druk om betekenis te geven aan de enorme hoeveelheden informatie die onze bedrijven genereren en opslaan. Om te kunnen beantwoorden aan de specifieke behoeften van elk bedrijf, moeten ze de juiste programmeertalen, modellen en tool sets - voor dataopslag, verwerking, analyse, visualisatie, enz. - combineren en configureren. Daar waar de meeste data science-platformen het aantal compatibele tools beperken, ontwikkelde Qrama, een imec spin-off die ook opgenomen is in het imec.istart-programma, het Tengu-platform: een generisch open-source platform dat professionals volledige vrijheid geeft om de beste tools te kiezen voor elk van hun data science-projecten. Tengu wordt momenteel gebruikt als onderdeel van imecs City of Things programma in Antwerpen.

Een nieuw, IoT-gedreven hoofdstuk in de datawetenschap

Tegen 2020 wordt verwacht dat er 44 Zettabytes aan data verzameld zullen zijn, wat een opslagcapaciteit vereist vergelijkbaar met 20% van de oppervlakte van Manhattan. Bovendien zal dit soort data door de komst van het Internet of Things nog veel sneller geproduceerd en verzameld worden en uit veel meer verschillende formaten bestaan. Naast de traditionele tekstuele en audiovisuele data zullen er ook nieuwe soorten data zijn, verzameld door allerlei sensoren, radars en andere (slimme) apparaten.

Betekenis geven aan al deze data is essentieel voor het succes van onze bedrijven. Om dat te kunnen doen, moeten datawetenschappers gegevens vanuit verschillende bronnen kunnen combineren en verkennen, modellen coderen en opbouwen om die gegevens in te zetten, deze modellen in productie toepassen en de resultaten op een begrijpelijke manier presenteren, via een rapport of een webapplicatie. Data science-platformen bieden een generische omgeving waarbinnen al de nodige tools gemakkelijk en automatisch gecombineerd kunnen worden zonder dat deze handmatig geconfigureerd moeten worden. Zo kunnen datawetenschappers zich meer op business intelligence concentreren en minder op het programmeren.

Geen kant-en-klare data science-oplossing

Het Internet of Things heeft geleid tot een sterke stijging in het aantal instrumenten voor datawetenschap, waardoor het moeilijk wordt voor platformaanbieders om de constante veranderingen en innovaties bij te houden. Daarom kiezen de meeste bedrijven ervoor om het aantal compatibele tools te beperken, meestal ten voordele van hun eigen oplossingen.

Elk data science-project vereist echter een unieke combinatie van tools voor elke stap van het proces – wat niet altijd mogelijk is met de huidige state-of-the-art-platformen. Daarom ontwikkelden Thomas Vanhove en Gregory Van Seghbroeck het Tengu-platform: een flexibel platform om snel Big Data-omgevingen op te zetten, te configureren en te beheren.

Jouw canvas voor datawetenschap

Tengu is een digitale werkomgeving voor Big Data-projecten. Het biedt een generisch en flexibel kader dat datawetenschappers toelaat om verschillende software te combineren en ermee te experimenteren zonder dat alles handmatig geconfigureerd of opgestart moet worden. Dit betekent een enorme tijdbesparing, aangezien dit proces soms zelfs enkele maanden in beslag kon nemen. Het Tengu-platform biedt zowel batch- als streamgeoriënteerde dataverwerking aan om in te spelen op de vereisten van het Internet of Things.

Medeoprichter Thomas Vanhove legt uit: "Traditionele datawetenschap was gebaseerd op grote datasets die allemaal tegelijk werden verzameld en zogenaamde batchanalyse vereisten. Met het Internet of Things is de focus verschoven van het analyseren van grote hoeveelheden data naar de mogelijkheid om verschillende types data te kunnen verwerken terwijl ze geproduceerd worden – wat we streamanalyse noemen. Aangezien batchanalyse nog steeds standaard is voor ongeveer 80% van de markt, zijn de meeste data science-platformen nog niet klaar om streamanalyse te ondersteunen. Dat is waar Tengu een echt verschil kan maken."

Het platform is compatibel met alle Big Data-verwerking, -opslag en -cloudcomponenten en is daardoor dus leveranciersafhankelijk. "Onze gebruikers kunnen vrij gebruik maken van om het even welke bedrijfsspecifieke of open-source oplossing die ze willen. Nadat de werkomgeving gebouwd is, kan die bovendien onafhankelijk van ons platform werken," benadrukt medeoprichter Gregory Van Seghbroeck.

Van academische nood tot bedrijfsoplossing

In 2013 was Thomas Vanhove promovendus aan imec - UGent, waar hij voor zijn onderzoek moest experimenteren met verschillende data science-tools. Al snel realiseerde hij zich dat de beschikbare platformen niet de flexibiliteit konden bieden die nodig was voor zijn werk, waarop hij besloot zijn eigen data science-platform te ontwikkelen.

Toen meer en meer collega's zijn platform begonnen te gebruiken voor hun eigen onderzoek, begon Thomas te bekijken wat de commerciële mogelijkheden van zijn idee waren. In 2015 nam hij deel aan de Opportunity Recognition Workshop (ORW) van imec, een driedaagse workshop voor onderzoekers en doctoraatstudenten om het potentieel van hun werk te bekijken vanuit een zakelijk en economisch perspectief. Nadat ze een businesscoach toegewezen kregen om hun idee verder te ontwikkelen, lanceerden Thomas Vanhove en Gregory Van Seghbroeck hun [start-up Qrama](#) om het Tengu-platform te commercialiseren. In mei 2016 werden ze vervolgens aanvaard voor het [imec.istart Business Incubation Programma](#)

Als spin-off van imec – UGent, kon Qrama van beide organisaties wereldwijd de exclusieve licentie verkrijgen voor de code achter het Tengu-platform, dat nu door imec gebruikt wordt in de context van het City of Things (CoT) programma.

Philip Leroux, Operational Lead van het City of Things Data Team van IDLab – Gent, benadrukt de waarde van Qrama's innovatieve technologie: "Het imec City of Things project in Antwerpen wil een open en toekomstgericht lab zijn voor smart city living en technologie. Door het Tengu-platform te gebruiken als de kern van ons CoT R&D Data Platform wordt het opzetten en onderhouden van de vereiste IoT datacomponenten veel eenvoudiger. Tegelijkertijd maakt Tengu het mogelijk om ons CoT R&D Data Platform uit te breiden voor toekomstige use cases. Met Tengu blijft ons platform erg flexibel en open voor nieuwe (big data) technologieën en server configuraties."

Een ondersteunend netwerk: de verborgen parel van imec.istart

Beide oprichters prijzen de steun van het imec.istart-programma bij het opschalen van hun bedrijf: "Het is een enorme voorsprong om in zo'n vroeg stadium al toegang te kunnen krijgen tot 50.000 euro. Dankzij die financiële ondersteuning konden we twee fulltime ontwikkelaars aanwerven, minder dan één jaar na de oprichting van ons bedrijf," aldus Thomas Vanhove.

De workshops van imec.istart hebben volgens beide ondernemers ook grote toegevoegde waarde, aangezien die hen hielpen om kennis te maken met alle commerciële, juridische en financiële aspecten van het starten en sturen van een bedrijf. "We zijn allebei technische mensen, dus we waren helemaal niet op de hoogte van wat het allemaal inhoudt om een zaak te runnen," erkent Gregory Van Seghbroeck.

Volgens de twee innovators is de heilige graal van imec.istart echter het robuuste netwerk van experts en partners, dat hen unieke opportuniteiten en exclusieve software- en zakelijke deals bood waardoor “de druk wat verminderde omdat we bepaalde taken konden uitbesteden aan imec.istart partners.” (Thomas Vanhove).

Volgende stappen

Nu ze al twee betalende klanten hebben binnengehaald en geselecteerd werden als het data science-platform voor het [City of Things-programma van imec](#), is het Qrama's doel om tegen eind juni 2017 tien nieuwe Europese klanten te overtuigen vooraleer ze in 2018 uitbreiden naar de Verenigde Staten. Ze hebben al verschillende conferenties in de VS bijgewoond om de markt en potentiële leads te leren kennen. “We richten ons vooral op adviesbureaus en grote bedrijven met eigen data science-teams,” licht Gregory Van Seghbroeck toe.

Doordat ze ook aanvaard werden voor een innovatieproject van VLAIO – het Vlaamse Agentschap voor Innovatie en Ondernemerschap – zoekt de start-up momenteel nog twee fulltime ontwikkelaars als extra ondersteuning. Om hun internationaliseringsplannen en het innovatieprogramma succesvol te kunnen uitvoeren en om hun platform verder te ontwikkelen, zoekt het bedrijf ook extra financiering.



Biografie Thomas Vanhove

Thomas Vanhove is medeoprichter en CEO bij Qrama. Hij behaalde een master Computerwetenschappen aan de UGent. In augustus 2012 startte hij zijn doctoraat bij de IBCN-onderzoeksgroep (nu IDLab) rond datamanagementoplossingen in cloudomgevingen. Tijdens zijn doctoraat ontwikkelde hij een eerste versie van het Tengu-platform om gemakkelijker te kunnen experimenteren met Big Data-structuren.

Biografie Gregory Van Seghbroeck

Gregory Van Seghbroeck is medeoprichter en CTO bij Qrama. Hij behaalde een doctoraat in Computerwetenschappen aan de UGent, waar hij momenteel werkt als postdoctoraal onderzoeker en betrokken is bij verschillende nationale en Europese projecten. Zijn werk werd gepubliceerd in tal van internationale tijdschriften en conferentieverslagen.

