



OQUILA Data & Analytics
e-book

Van ware- tot lakehouse: hoe kies je de juiste data-oplossing?



DE WERELD VERANDERT SNEL

IS UW DATAWAREHOUSE GEBOUWD VOOR DE WERELD VAN MORGEN?



Databronnen en datagerelateerde vragen evolueren continu. Een goede data-oplossing of -architectuur groeit dan ook mee met uw organisatie en voortschrijdende inzichten. Maar welke oplossingen bestaan er precies? En welke is het meest geschikt voor uw uitdagingen van vandaag én die van morgen? Verken in deze e-book de meest courante oplossingen en belangrijkste principes.



NAAR EEN 'SINGLE VERSION OF TRUTH'

Data-driven beslissingen beginnen vaak met een aantal eenvoudige rapporten. Tools zoals **Microsoft Power BI** connecteren daarvoor rechtstreeks met backendsystemen.

Naarmate er meer vragen komen en er dus meer informatie nodig is, zijn er alsmear meer systemen (ERP, CRM, Finance, ...) betrokken. Klantengegevens kunnen bijvoorbeeld in meerdere tools zitten, vaak met eigen coderingen, eigenschappen, ...

De informatie kan bovendien complementair of net overlappend zijn. In het laatste geval is het niet altijd duidelijk welke versie het meest accuraat is.

Met Power BI kunt u informatie transformen en mergen. Volstaat dat? Is dit in ieder rapport consistent toegepast? Wat als de regels wijzigen en transformaties gefragmenteerd zitten over meerdere rapporten of systemen? Een gecentraliseerde aanpak kan op termijn heel wat voordelen opleveren.



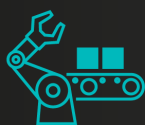


DATAKWALITEIT

MAAK BETROUWBARE BESLISSINGEN

Naast een gefragmenteerde datacollectie en -transformatie komen incorrecte, incomplete of dubbele gegevens vaak voor. Door menselijke fouten, historische fouten of meer structurele oorzaken. Het gevolg? Slechte beslissingen, geërgerde klanten, een gehavende reputatie of zelfs verminderde omzet.

Daarom is het datakwaliteit zo cruciaal bij het bepalen van een datastrategie op lange termijn en het kiezen van de juiste data-oplossing.



START KLEIN, MAAR **DENK GROOT**

Dankzij cloudtechnologie (waaronder **Microsoft Azure**) is het mogelijk om data intelligence-oplossingen en uw datastrategie gradueel uit te bouwen.

In de cloud kunt u immers stap voor stap schakelen, zonder grote investeringen. Wanneer u merkt dat de limieten van uw huidige rapportering bereikt zijn, stapt u gewoon over naar een structurele oplossing. Heel 'agile'.

Zo boekt u snel resultaat zonder het grotere plaatje uit het oog te verliezen. Bovendien bespaart het u heel wat extra werk nadien.



HET DATAWAREHOUSE ALS DATA-OPLOSSING

Data warehouses spelen al geruime tijd een belangrijke rol bij het ondersteunen van businessprocessen en rapportering.

Een datawarehouse helpt om getransformeerde, gefilterde en gestructureerde data – vaak transactionele data – te bewaren in een vaste, relationele datastructuur. Aan de hand van tools zoals Microsoft Power BI data zet u die data om in in visueel aantrekkelijke dashboards.

Een datawarehouse bestaat uit een set van technologieën, van ETL-componenten – zoals Azure Data Factory – tot databases – zoals Azure SQL.

android...
license.html
...ing source files with shaisums...
tools-linux-4333796.zip ... Passed
id-sdk.sh ... Passed
id-sdk.csh ... Passed
...nf ...

22651 saikiran
902 saikiran
472 root
21853 saikiran
380 root
1 root
231 root
10981 root
568 saikiran

DATAWAREHOUSE

VOORDELEN EN BEPERKINGEN

Een datawarehouse zorgt ervoor dat data uit verschillende systemen klaargestoomd worden voor de business. Er zijn echter ook een aantal beperkingen:

- Data warehouses kunnen heel goed om met gestructureerde data aan. De hoeveelheid semi-gestructureerde en ongestructureerde data – zoals foto's en video's – die organisaties moeten verwerken neemt echter steeds verder toe.
- Niet alleen het aantal mogelijke datatypes, ook de nood aan opslagruimte is enorm toegenomen. Denk maar aan video. In een klassiek datawarehouse kunnen de kosten daardoor enorm oplopen.
- Een datawarehouse heeft een relationeel schema. Nieuwe datastromen of -systemen toevoegen is hierdoor minder evident. Hou daar rekening mee bij het wegschrijven van data.
- Het schema is opgezet volgens de vragen die vandaag gesteld worden. Nieuwe inzichten en toekomstige noden kunnen daardoor een drastische impact hebben op bestaande ETL pipelines, rapporten, ...
- De vraag naar data groeit, ook binnen organisaties. Zo hebben naast businessanalisten en report designers ook data scientists, AI-modellen en machine learning algoritmen een groeiende nood aan ruwe, ongeprepareerde data.

DATE LAKE

BEPERKINGEN OMZEILEN ... OF NIET?

Om bovengenoemde beperkingen te omzeilen, maken heel wat organisaties gebruik van een data lake. In tegenstelling tot een data warehouse is een data lake namelijk niet onderhevig aan vaste structuren, en dus flexibeler en makkelijker uit te breiden. Een data lake laat toe om vrijwel alle ongestructureerde data weg te schrijven en uit te lezen wanneer nodig. Dat maakt deze oplossing een stuk toekomstgerichter, en de droom van elke data scientist.

Toch is een data lake niet in alle omstandigheden de ultieme oplossing. Zo zijn ook hier een aantal beperkingen:

- De datakwaliteit is niet gewaarborgd, en er is meer kans op inconsistenties en ruis. Dat leidt tot verkeerde cijfers en dus foute conclusies en beslissingen.
- Voor business users is een data lake nagenoeg onbruikbaar, aangezien alle data getransformeerd moet worden.
- Een data lake biedt geen goede ondersteuning voor transacties.
- Data afschermen of enkel ter beschikking stellen van selecte personen is complex.

Data lakehouse: het beste van twee werelden

Gelukkig bestaat er sinds kort ook zoiets als een data lakehouse. Deze oplossing combineert de voordelen van een data lake en warehouse in één architectuur of oplossing. Zo biedt een lakehouse zowel de flexibiliteit en diversiteit van een data lake als de consistentie en kwaliteit van een data warehouse. Kwaliteitsmonitoring zorgt er bovendien voor dat data steeds compleet en correct is.

	Data Lake	Data Warehouse
Data type	Structured, semi-structured, unstructured. Relational, non-relational	Structured, Relational
Schema	Schema on read	Schema on write
Format	Raw, unfiltered	Processed, vetted
Sources	Big data, IoT, social media, streaming data	Application, business, transactional data, batch reporting
Scalability	Easy to scale at a low cost	Difficult and expensive to scale
Users	Data scientists, data engineers	Data warehouse professionals, business analysts
Use Cases	Machine learning, predictive analytics, real-time analytics	Core reporting, BI

	✓ Lakehouse
Data type	Structured, semi-structured, unstructured Relational, non-relational
Schema	Schema on read, schema on write
Format	Raw, unfiltered, processed, curated, delta format files
Sources	Big data, IoT, social media, streaming data, application, business, transactional data, batch reporting
Scalability	Easy to scale at a low cost
Users	Business analysts, data engineers, data scientists
Use Cases	Core reporting, BI, machine learning, predictive analytics

In een data lakehouse zijn opslag en rekenkracht losgekoppeld van elkaar. Dat maakt deze oplossing zeer schaalbaar en kostenefficiënt, ook bij grote hoeveelheden data. Rekenkracht kan op die manier immers ingezet worden wanneer dat nodig is. Kortom, met een data lakehouse kunnen zowel data scientists als businessanalisten en users optimaal gebruik maken van data.

Security van
devices, connecties
en cloud.

Schaalbaarheid
zodat u uw
data- en/of (I)
IoT-oplossingen
snel kunt schalen
naargelang uw
behoeften.

Compatibiliteit
met bestaande
(legacy) systemen
en software.

Pre-
geconfigureerde
data- en (I)IoT-
templates en
scenario's, zodat
u snel aan de slag
kunt.

Kostenefficiënt
dankzij het 'pay as
you go'-model.

DATA LAKEHOUSE

WAAR MOET U OP LETTEN?

- Ga waar mogelijk voor een **standaardoplossing** die voldoende gedocumenteerd is. Zo kunt u bepaalde taken of configuraties ook zelf doorvoeren en vermijdt u vendor lock-in.
- Kies voor een **'open' systeem** met een groot aanbod aan connectoren. Zo blijft uw dataplatform de centrale plaats van waaruit alle data, ongeacht hun herkomst, op een uniforme en gecontroleerde manier toegankelijk zijn.

” **Elke situatie is uniek, maar deze criteria helpen u alvast bij uw keuze.**

- **Vermijd rapporteringssystemen** die enkel een uitbreiding zijn op ERP- en CRM-systemen, om een centraal dataplatform te behouden.
- On-premise systemen of virtuele machines (on-prem of cloud) vergen **manueel onderhoud**. Bij het gebruik van PaaS- of SaaS-componenten valt die zorg weg.
- Een dataplatform moet instant **schaalbaar** zijn en kunnen **meegroeien** met uw organisatie. Zo zijn er bijvoorbeeld piekbelastingen bij de opstart van het project (full load), heeft u na de kantooruren wellicht minder resources nodig, ... Een schaalbaar systeem zorgt er ook voor dat u niet vanaf dag één betaalt voor alle mogelijke resources, maar stapsgewijs kunt uitbreiden.

BOUW MET OQUILA

START KLEIN, MAAR DENK GROOT



OQuila maakt gebruik van **100% Azure-componenten** (PaaS en SaaS). We combineren de voordelen van een data warehouse en data lake door middel van Azure Synapse en standaard OQuila-configuraties.

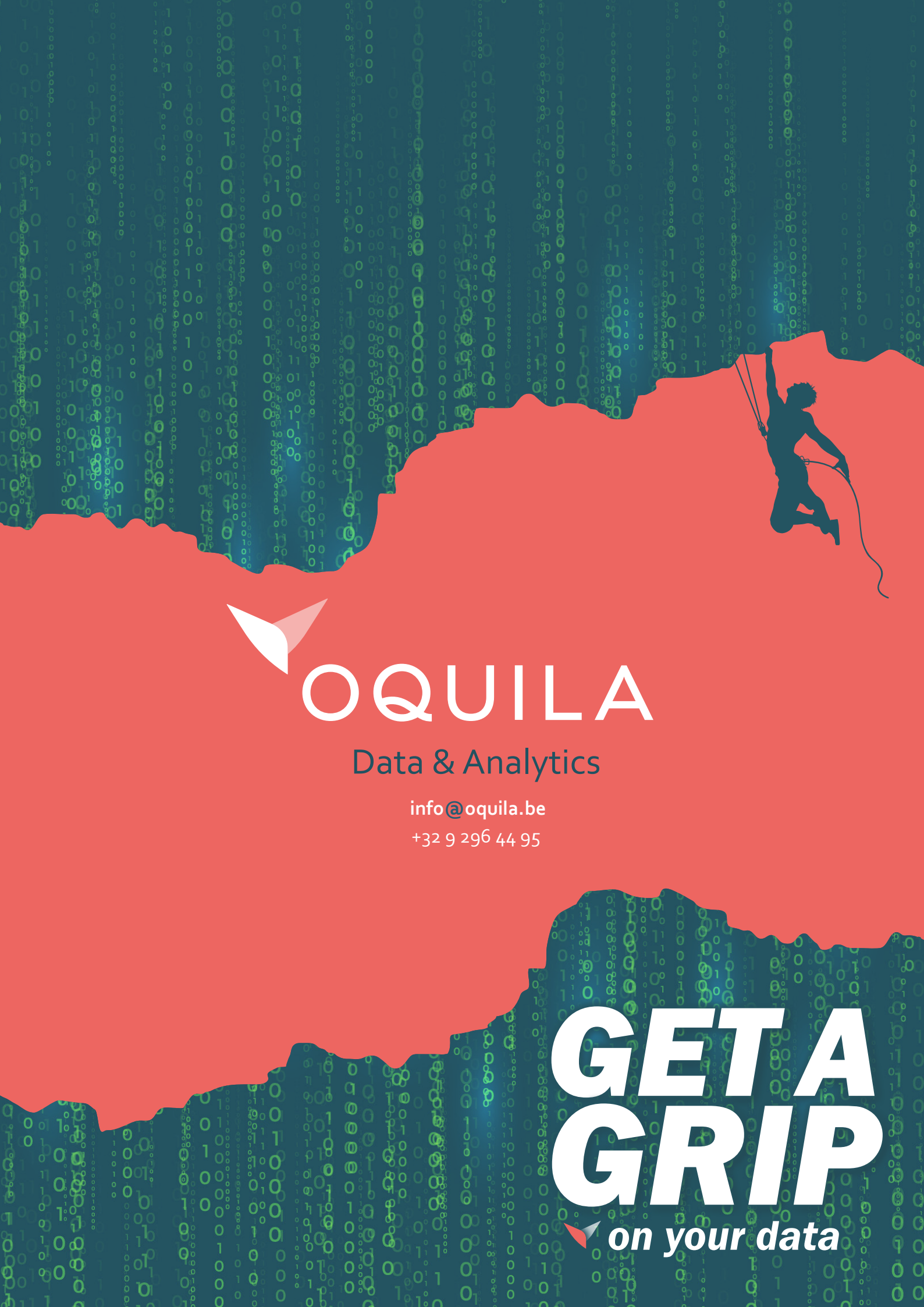
Via **standaardconnectoren** kan **brondata** uit **nagenoeg elk systeem** worden **gehaald**, en dat zowel on-premise als in de cloud. Zo kunnen zowel data scientists, ervaren businessanalisten als self-service BI-gebruikers ten volle de kracht van data benutten.

GET IN TOUCH!

Grote objectieven rond data-digitalisering? Bij OQuila begint alles begint met een luisterend oor en een gezonde portie nieuwsgierigheid.

Contacteer ons op:

info@oquila.be
[+32 9 296 44 95](tel:+3292964495)



OQUILA

Data & Analytics

info@oquila.be

+32 9 296 44 95

GET A GRIP



on your data