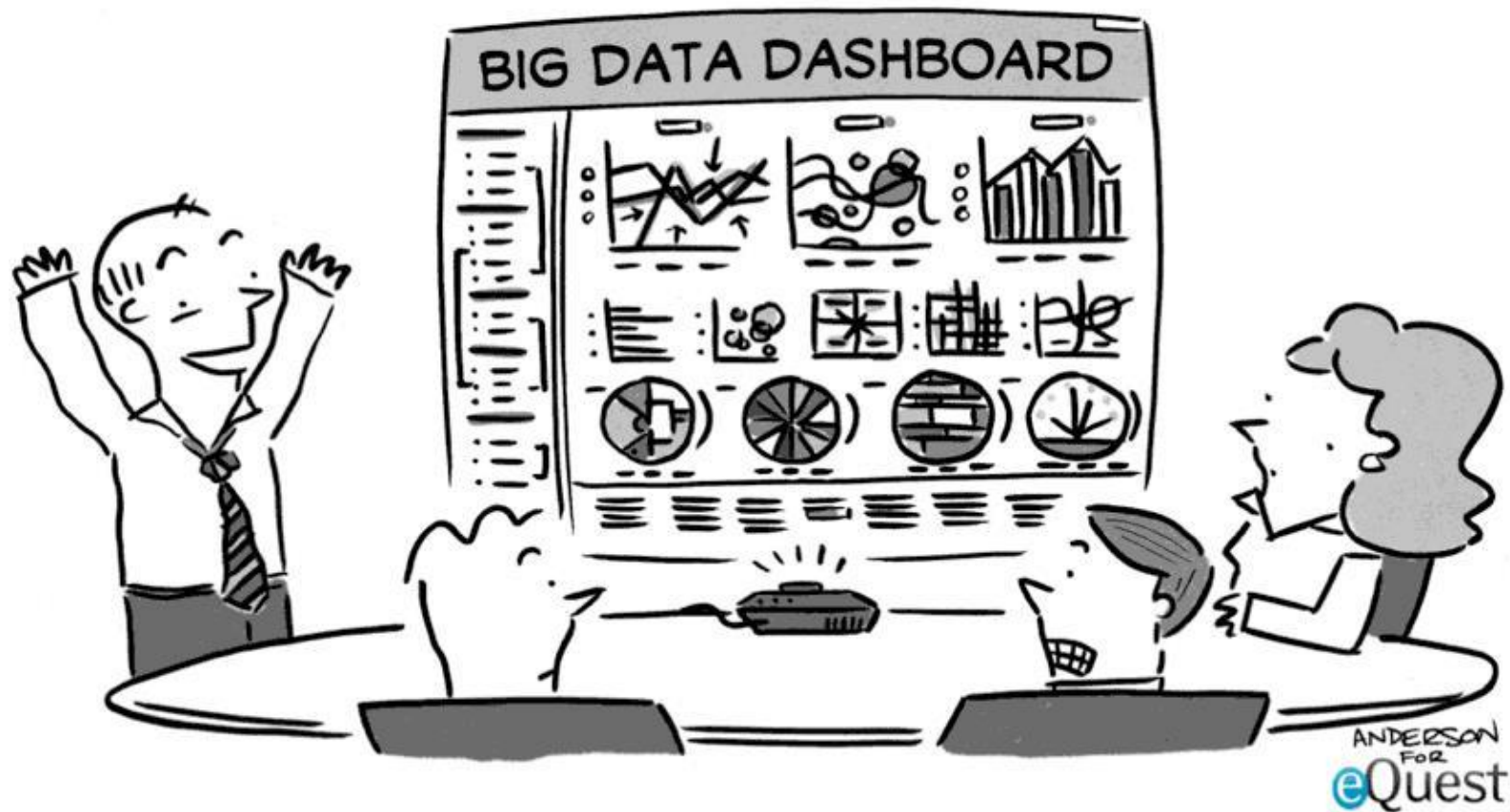


Granulair risico inzicht op basis van Big Data

Koen Rutten



Vlaanderen
is mobiliteit &
openbare werken



"After careful consideration of all 437 charts, graphs, and metrics, I've decided to throw up my hands, hit the liquor store, and get snocked. Who's with me?!"

Agenda

- INLEIDING
 - Doelstelling
 - Diensten
- BUSINESS CASE
 - Algemeen
 - Uitdaging
 - Resultaat & toekomst
- DATA PLATFORM
 - Technologie
 - Evolutie: ad-hoc > always-on > serverless



Vlaanderen
is mobiliteit &
openbare werken

INLEIDING

- Doelstelling
- Wat biedt het Datalab aan



Doelstelling

INNOVATIEF DATAONDERZOEK

Het Datalab van MOW biedt expertise en innovatieve oplossingen aan omtrent data science en analytics.

De focus van het team ligt op het onderzoeken van data om inzichten en patronen te ontdekken, voorspellingen te maken en aanbevelingen te genereren, met behulp van geavanceerde technieken zoals big data, artificial intelligence en machine learning.

Experten ter ondersteuning van het departement en de samenwerkingsovereenkomsten met agentschappen, in het kader van:

Pilootprojecten

Advies ter ondersteuning

Algoritmes en programmeren

Data storage

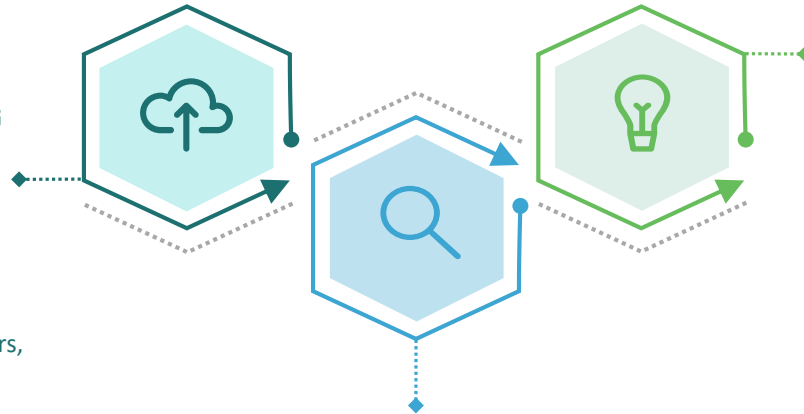
Data analytics

AI en Machine Learning expertise

Wat biedt het datalab aan?

3. PLATFORM & ONDERSTEUNING

- Data captatie
- Borging van code & data
- Beheer van data op public Cloud platform
- Platform aanbieden als sandbox omgeving (pilotomgeving, clusters, tools, etc.)



1. INZICHTEN EN MODELLEN CREËREN VIA PROJECTAANPAK

- Basis- en geavanceerde algoritmen toepassen
- Het creëren en toepassen van relevante technieken om inzicht en business impact te creëren
- Samenwerken met de experts

2. KENNIS EN EXPERTISE

- Technische vertaalslag tussen externe partijen en interne stakeholders voor vereisten van data science/analyse werk
- Ad-hoc data science analyses
- Ondersteuning in aanrijking van technologische mogelijkheden



Vlaanderen
is mobiliteit &
openbare werken

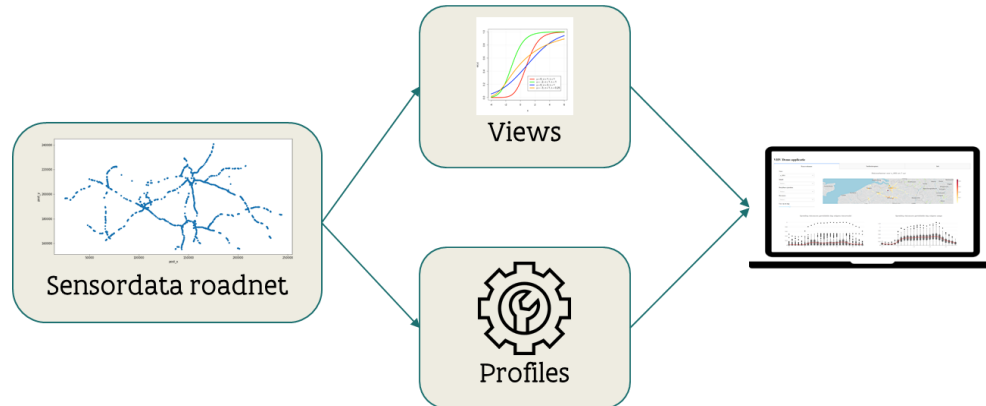
BUSINESS CASE

- Algemeen
- Uitdaging
- Resultaat & toekomst

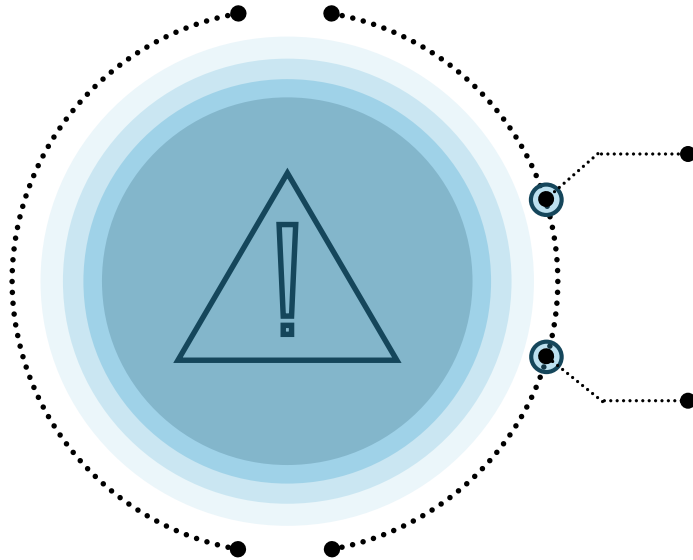


Algemeen

- Rapporteren over gevaarlijke plekken op het Belgisch netwerk (EU - Network Safety Ranking)
- Evalueer risico, veiligheid en kans op ongevallen om beleidsmakers te adviseren
- Reageer op ad-hoc vragen



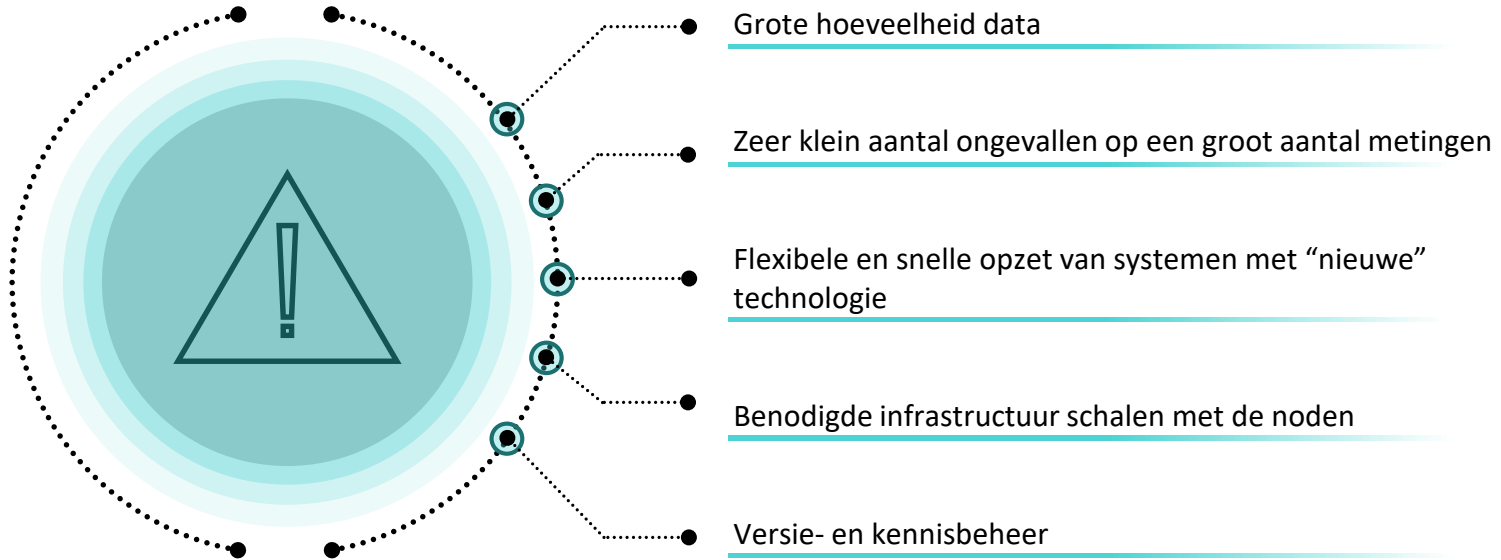
Stakeholder uitdagingen



Conclusies trekken op basis van jaarlijkse aggregaten
(vanwege hoeveelheid gegevens)

Risico en veiligheid gebaseerd op ongevallen die gebeurd zijn
→ Blinde vlek voor risico als er (nog) geen ongevallen zijn

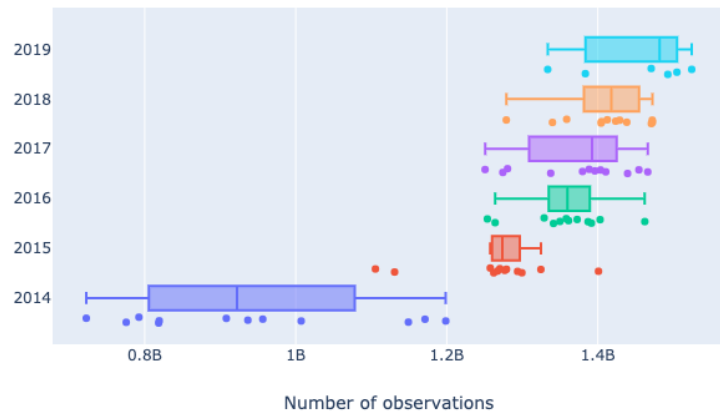
Technische uitdagingen



Data uitdagingen

- Raw measurement data
 - 2014 – juni 2019
 - 3,94 TB ruwe CSV
 - 1,42 TB Parquet bestanden
 - ± 1,3 miljard observaties per maand
 - Gemiddeld aantal observaties per maand stijgt elk jaar

Number of observations per month

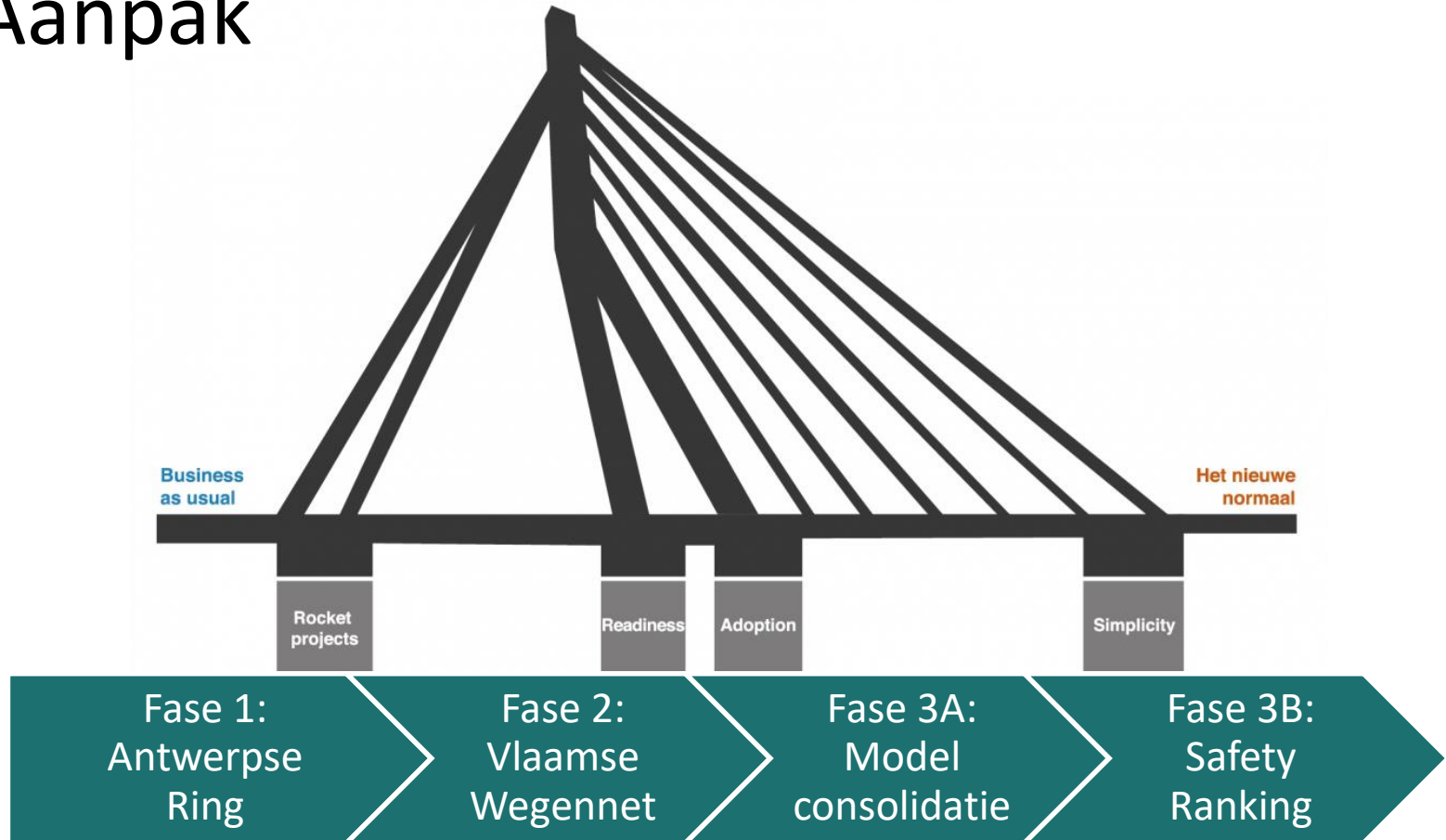


- 2377 letselongevallen (FEDPOL 2014-2018)
- 92976 incidenten (Vlaams Verkeerscentrum 2014-2018)

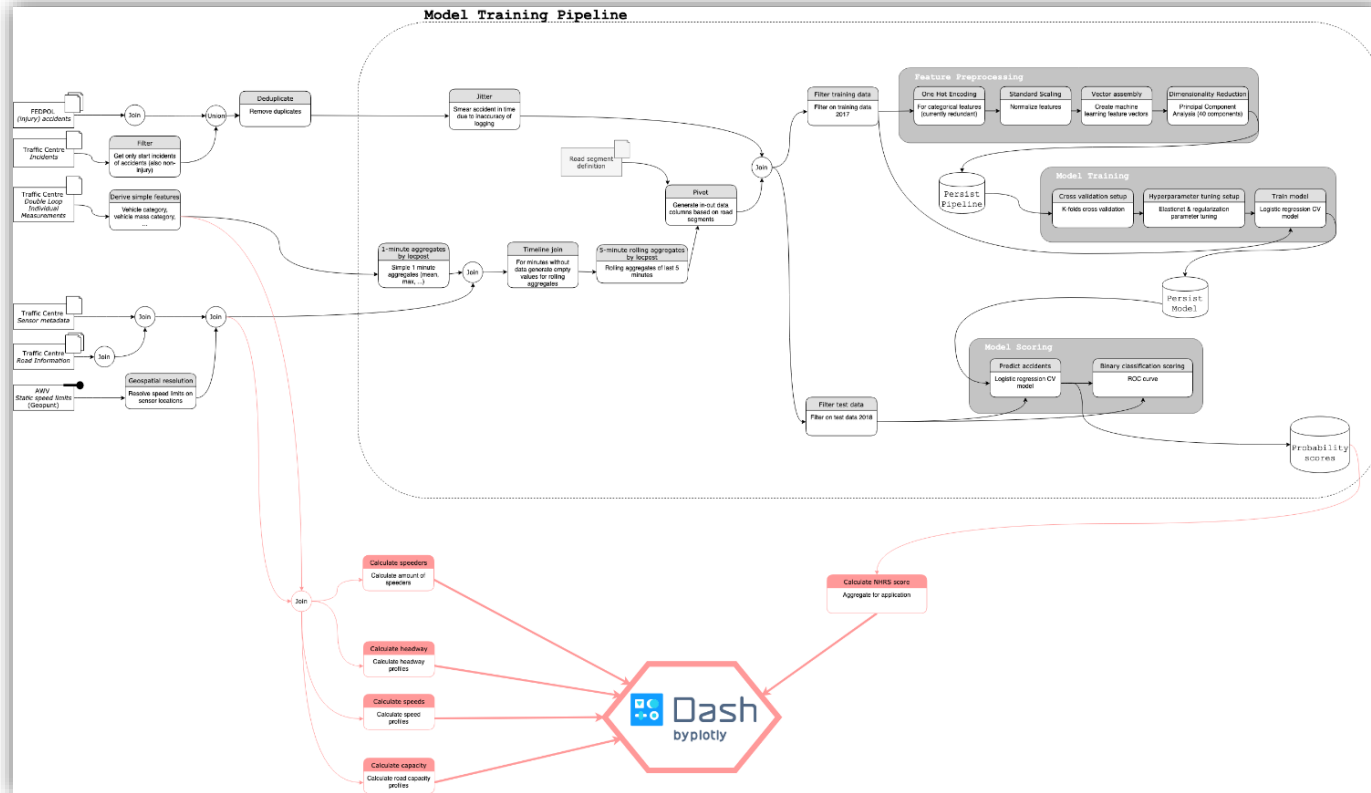


"I think you'll find that mine is bigger..."

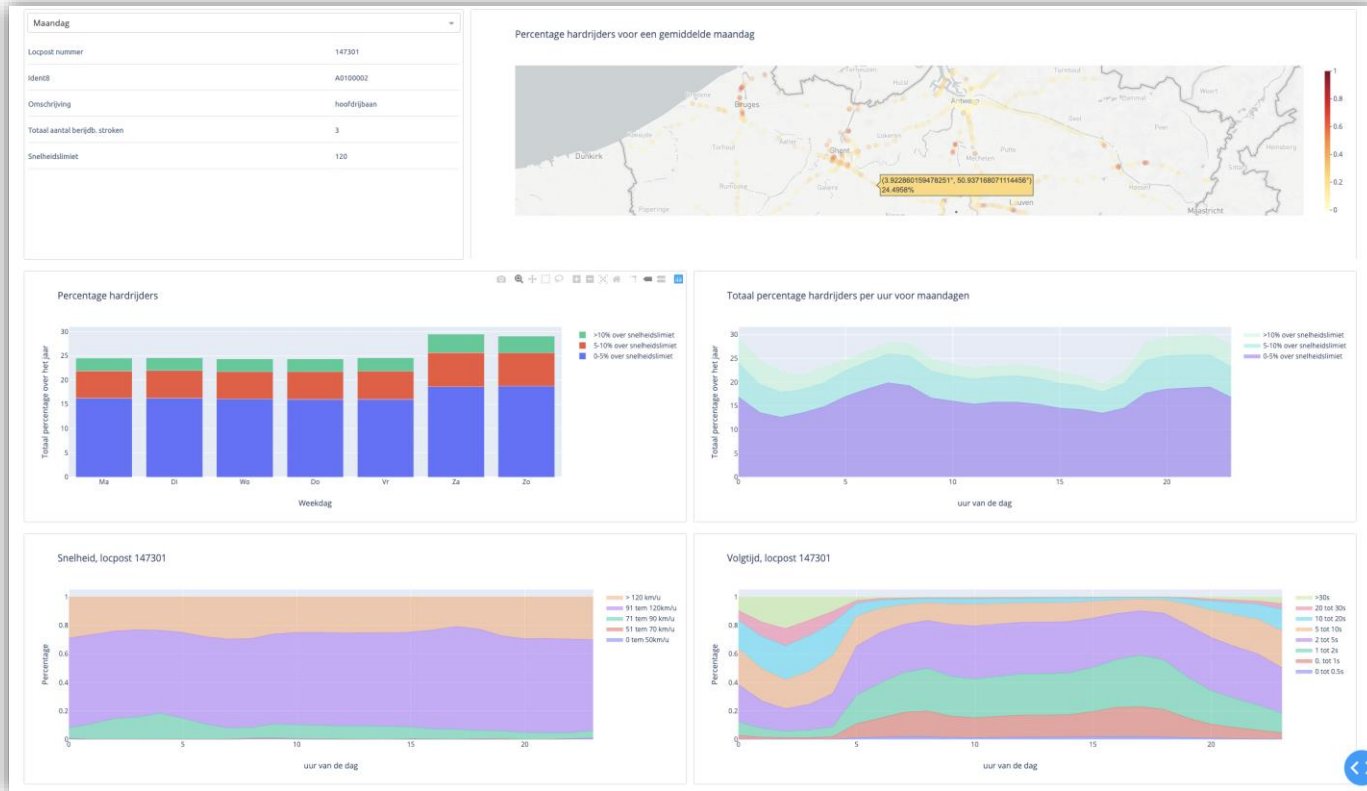
Aanpak



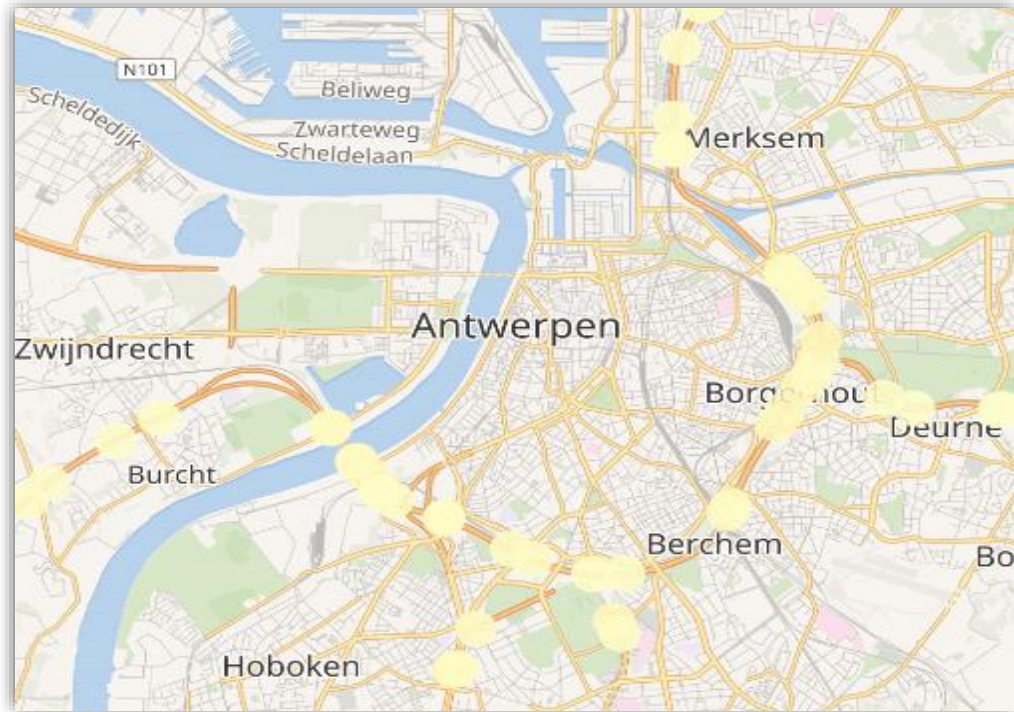
Opbouw van complexiteit



Demo (1)



Demo (2)



** gemiddelde dag op de R1. **



Vlaanderen

is mobiliteit &
openbare werken

DATA PLATFORM

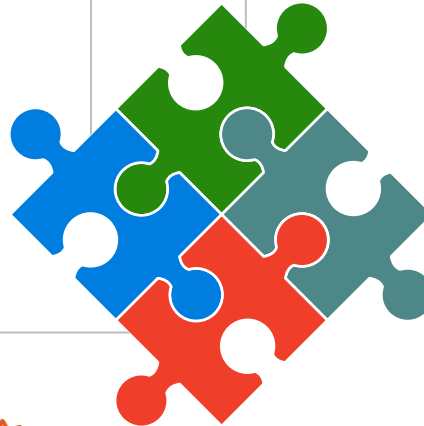
- Technologie
- Evolutie: ad-hoc → always-on → serverless

Welkom in onze Cloud

Azure Cloud omgeving in beheer van het Datalab

DATA STORAGE

Alle data van projecten wordt bewaard in de Cloud omgeving als primaire bron



INFRASTRUCTURE-AS-CODE

De infrastructuur is opgezet in de vorm van infrastructure-as-code en zit in ons versiebeheer.



AZURE DATABRICKS

Notebook en job omgeving om scripts in Python, R & Scala uit te voeren.



AZURE DEVOPS

Versie- & kennisbeheer van code via Git in Azure DevOps.



Platform evolutie



AD-HOC LOKAAL

- Installatierechten
- Bepaalde compute
- Grote hoeveelheden data



CLOUD IAAS/PAAS

- Scaling mogelijk maar moeilijker
- Always-on: “duur”
- Meer flexibiliteit
- Upgrades soms moeilijker



CLOUD SAAS

- Scaling zeer flexibel
- Compute enkel wanneer nodig
- Upgrades zeer flexibel

Platform kosten



AD-HOC LOKAAL

- Niet feasible te veel data
- Installaties zorgen voor vertraging



CLOUD IAAS/PAAS

- Grootste kost: infrastructuur
- Tweede kost: data storage



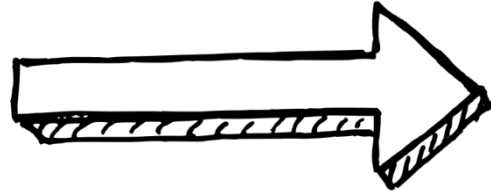
CLOUD SAAS

- Grootste kost: data storage
- Compute kosten gereduceerd met 80%-90%

Platform optimalisaties



CLOUD IAAS/PAAS



- Data storage format
- Partitionering
- Data storage type
- Versioned code



CLOUD SAAS

Vragen?

KOEN.RUTTEN@MOW.VLAANDEREN.BE

