

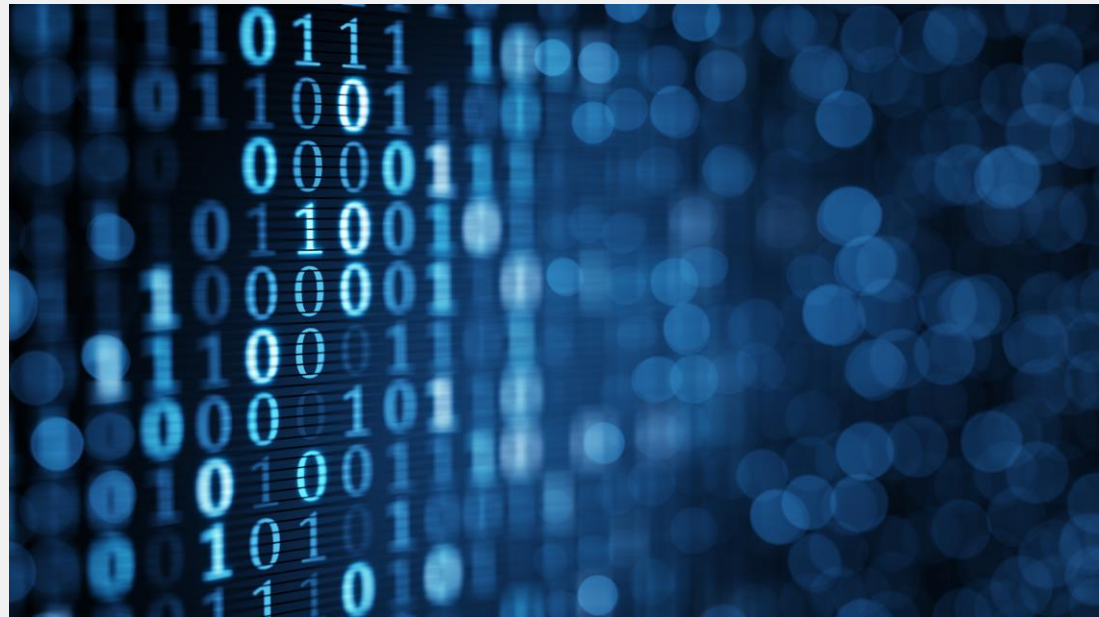
HTM



Big Data bij HTM

Datalab

29 januari 2020

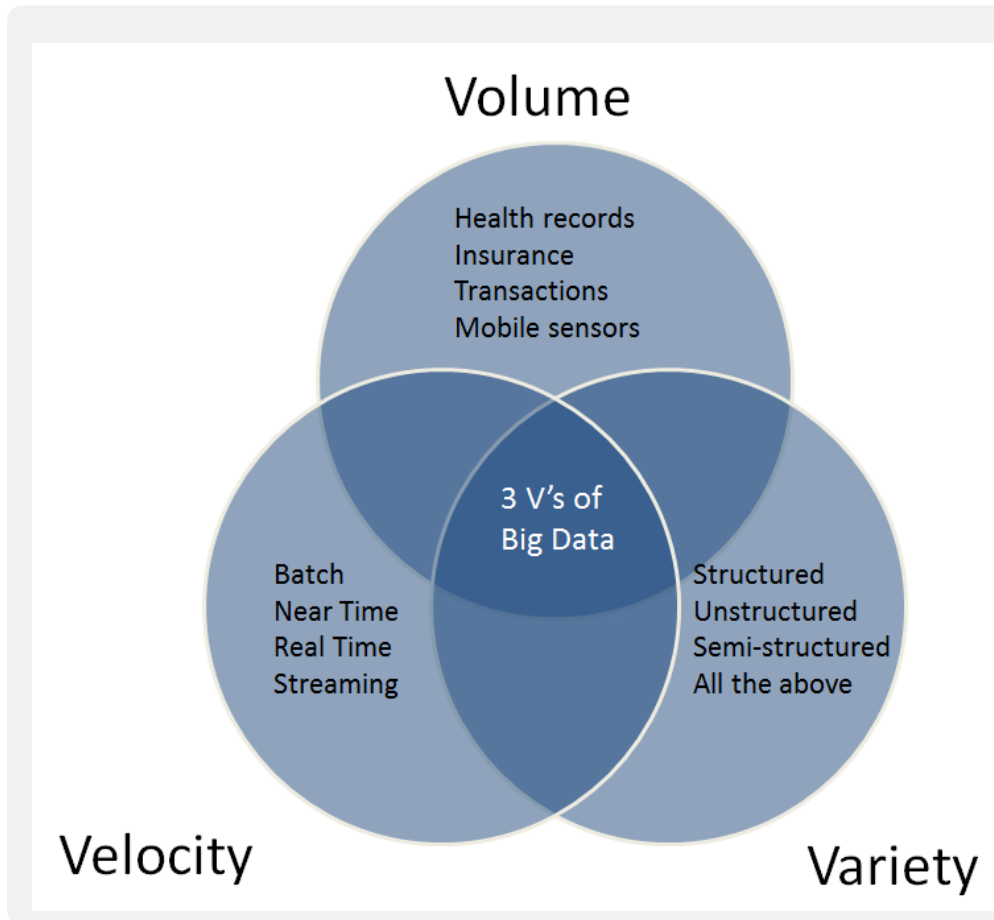


Inhoud

- Big data?
- Aanleidingen en uitdagingen bij HTM
- Datalab – aanpak methode
- Datalab – aanpak techniek
- Use cases
- Toekomstbeeld



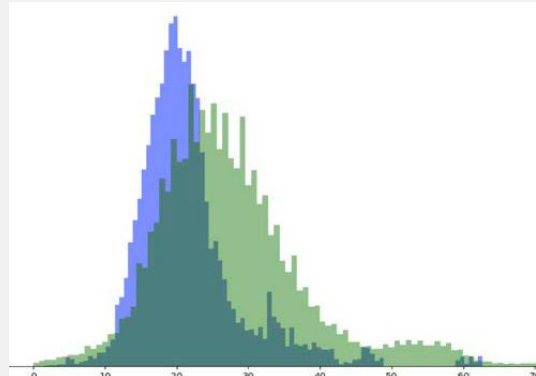
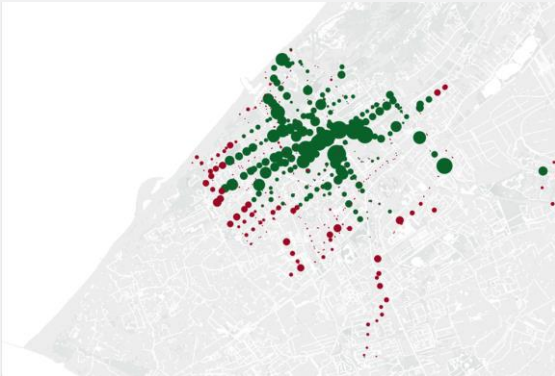
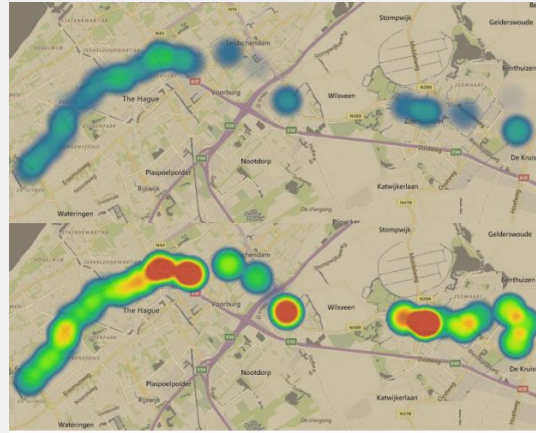
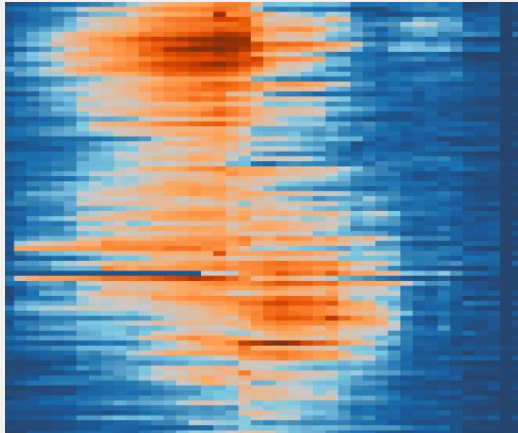
Big Data: hype?



- Concept is niet nieuw
- Nieuwe technieken noodzakelijk: Distributed Computing
- Internet of Things, sensing, Streaming, Time Series
- Kansen



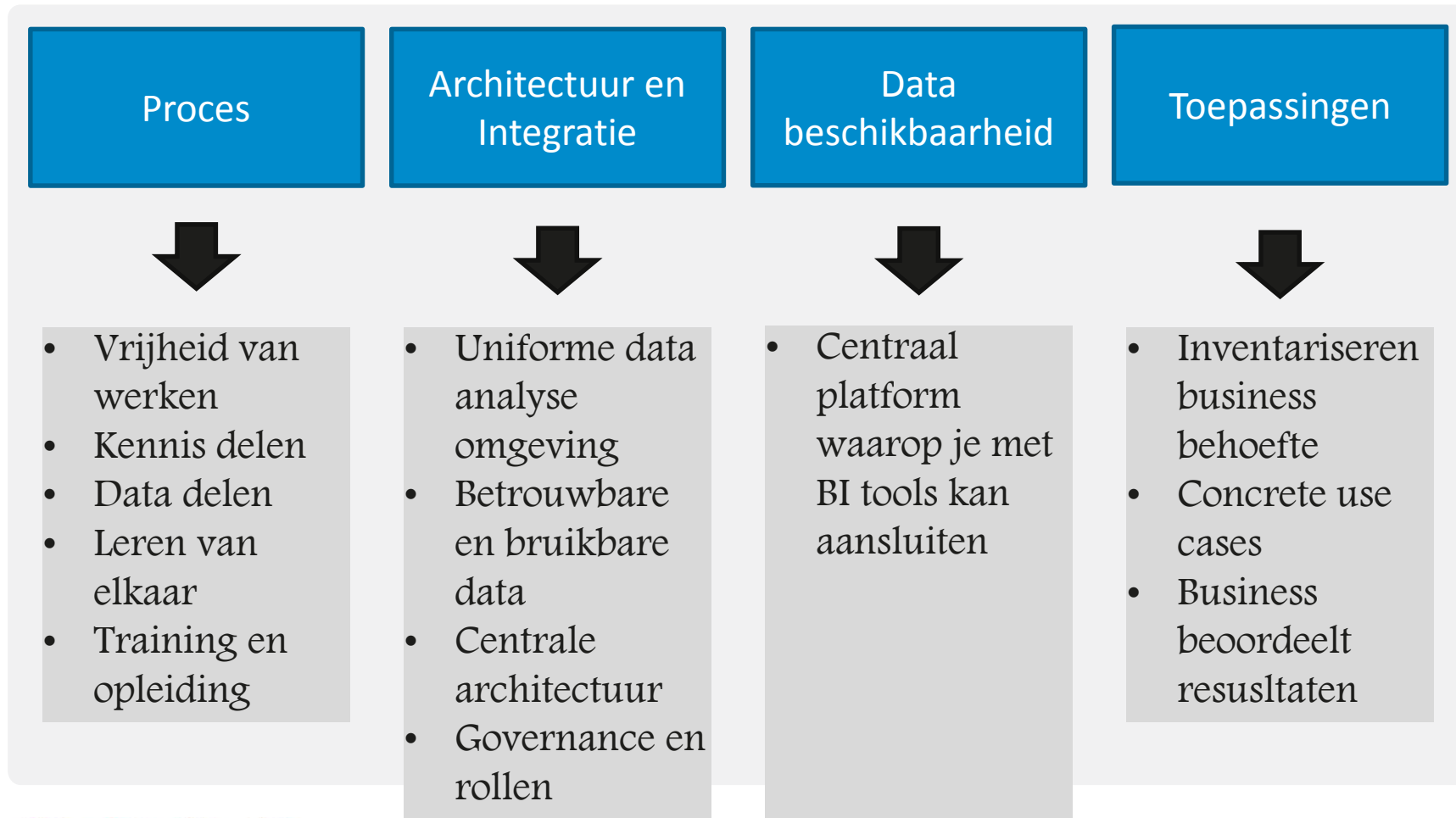
Big Data in het Openbaar vervoer



- Techniek
- Reizigers
- Tactische planning
- Finance
- IT (chipkaart, EBS)



Welke aanleidingen waren er bij HTM?



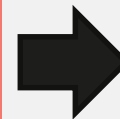
Er is wel een aantal uitdagingen op de weg (1/2)

De rol van HTM IT was niet altijd duidelijk



HTM IT neemt duidelijk rol als facilitator van het initiatief

Er is geen centrale analyse omgeving



In Q3/Q4 besuiten we welk platform geschikt is als analyse omgeving. Mogelijk speelt de KPMG Azure platform een rol.

Er is onvoldoende overzicht over het bestaan ende beschikbaarheid van verschillende databronnen



HTM IT neemt het voortouw in het maken van een data-catalogus waarmee je kunt navigeren door het HTM data-landschap

Onvoldoende tijd/middelen



We vragen het DT om hulp in tijd en middelen en automatiseren meer om de werkdruk te verlagen



Er is wel een aantal uitdagingen op de weg (2/2)

Onduidelijk wie eigenaar is van welke data



Richtlijnen Data Governance moet door HTM ICT worden geboden

We weten onvoldoende van elkaar wat we doen, wat we kunnen



DataLab lost dit van nature al op. Data-collega's krijgen een eigen Sharepoint omgeving.

Werken met data op detailniveau kan conflicteren met privacy wetgeving/
concurrentiegevoeligheid



Nauw samenwerken met privacy officer / security officer. Data platform moet voldoen aan hoge security standaarden

Externe data is moeilijk bruikbaar omdat een goede landingsplaats ontbreekt



Integratie van (externe) data moet een integraal onderdeel zijn van data-platform

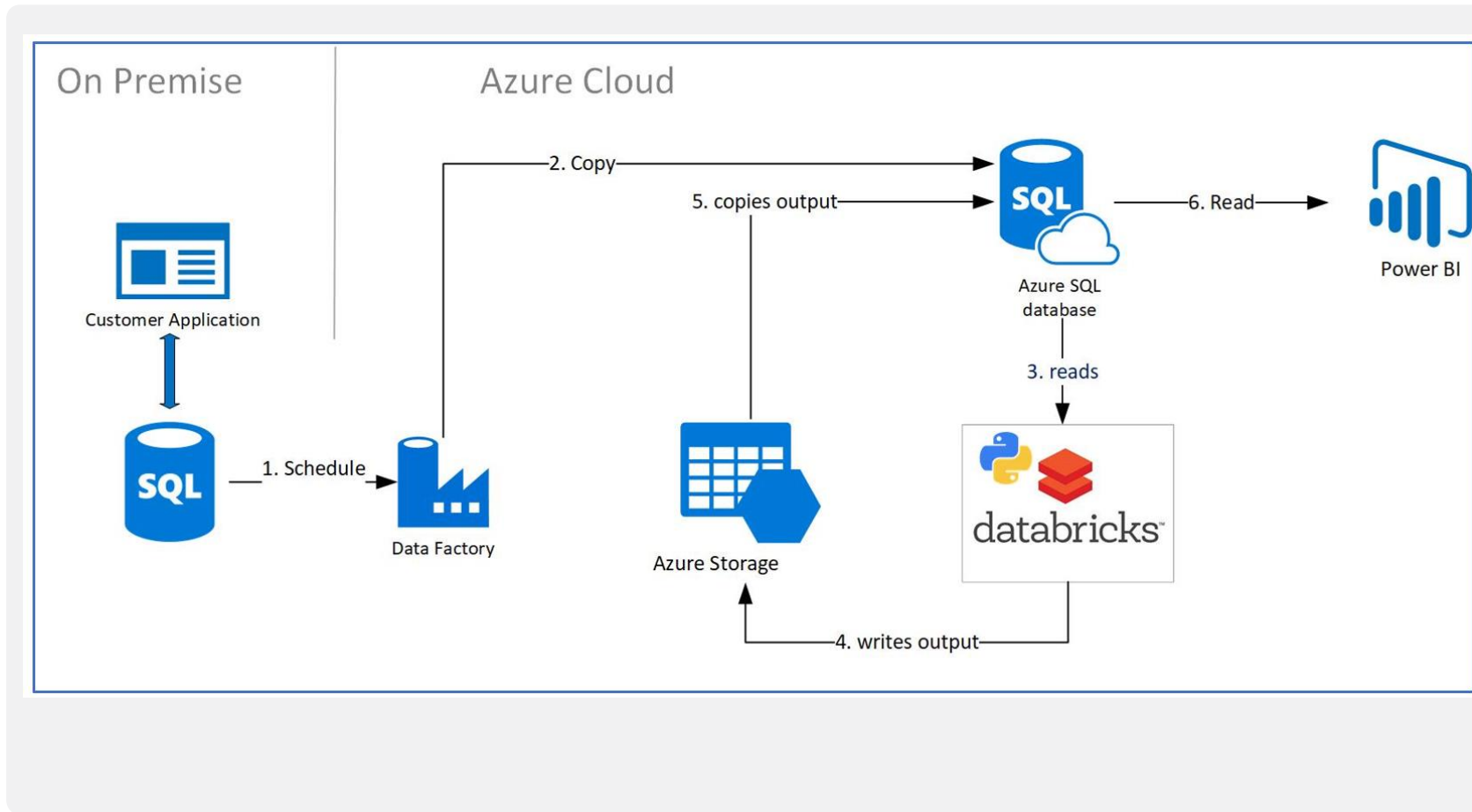


Wat doen de HTM datalab deelnemers?

- *Belangrijk mandaat: n% tijd voor de werkzaamheden*
- *Belangrijk principe: DL <> legacy BI*
- *Belangrijk principe: Agile – DevOps Lite*
- *Belangrijk principe: leertraject (reis, doel, kennis van buiten)*
- Scientists – Engineers – Stewards
-
- Data Mining
- ML: supervised, unsupervised (clustering)
- En vooral: meerwaarde bouwen dmv meervoudige bronnen



Aanpak – Platform



Use Cases

De relatie tussen haltepassages en conditie wielen/infra

- Koppelvlakken
- Wielslijtage
- Infra Status

Het aantonen van causaliteit tussen incidenten / verstoringen en operationele gevolgen/reizigersvertraging

- OV-Chipkaart
- Koppelvlakken
- Incident-mngmt

Het beheersen van tractie-volumes in het algemeen en van e-bus verbruik in het bijzonder

- Tractie-data
- E-bus stats
- Koppelvlakken



In de toekomst

- API catalogus
- Zo min mogelijk ‘snapshots’
- Veel meer IoT
- Eindproducten als volwaardig bron voor BI – LowCode

