

UPGRADE JOUW KENNIS OVER DATA MANAGEMENT

'Community Data management'; Bert, Willem, Ben, Taco, Wim, Eric,...

| | | | |
|--------------------------------|--|-----------------------------|---|
| sessie 1 24 okt 2023 | Data Governance | Frank Harland | Een data governance agenda opstellen en in de praktijk brengen. |
| | Data Security: Zero Trust Architecture | Taco de Haan (TFG Security) | Zero Trust Architecture (ZTA) stelt dat alle bronnen op een veilige manier toegankelijk zijn. Het is een data-centrische benadering die nodig is voor de toekomst van informatiebeveiliging. |
| sessie 2 30 Januari 2024 | Data Architectuur over het Hoe en Waarom | Ben Teunissen | De eisen aan de DA worden gesteld vanuit de eisen en wensen, die de organisatie aan zijn data stelt. De DA biedt inzicht in deze data: creatie, gebruik, samenhang en levenscyclus. |
| | Data kwaliteiten verhogen met eenvoudige hulpmiddelen | Bert Dingemans | Wil je de kwaliteit van data in je organisatie verhogen hoe pak je dat op een pragmatische manier aan. In deze sessie laat ik een eenvoudig en praktisch raamwerk zien waarmee je op basis van patronen met behulp van maatregelen data kwaliteit kunt verhogen. |
| 6 & 20 maart 2024 | Training Certified Data Management Professional (CDMP) | Ronald Baan Dama-NL | 2 daagse training om goed voorbereid te zijn om het internationale examen certificaat CDMP te kunnen doen. Examen wordt apart afgenomen via DAMA international. |
| sessie 3 14 mei 2024 | De evolutie van Data platforms; Van Warehouse naar Lake naar Lakehouse naar... | Eric Fickel | Waar we in het verleden onze data opsloegen in lokale relationele databases is het tegenwoordig gangbaar om onze data in de cloud op een zgn. data platform op te slaan. Makkelijker vindbaar en beschikbaar te maken voor gebruikers; data market places voor afnemers en aanbieders. Hiervoor zijn verschillende architectuur concepten in de tijd langsgekomen. Momenteel is het data Lake house in opmars. In deze presentatie gaan we kort langs de tijdlijn van oplossingen richting o.a. het data lakehouse en andere nieuwe concepten en de drivers achter de concepten. |
| | Naar een data architectuur voor het nieuwe data fundament van NS | Yorrick Tillemans NS | NS werkt al jaren aan de ambitie om data gedreven te zijn. Op gebied van data architectuur knelt het daar momenteel. De centrale NS data architectuur uit 2015 sluit steeds minder aan bij de vergrote schaal, nieuwe ontwikkelingen en groeiende data ambities. NS heeft daarom gekozen om een nieuwe data architectuur op te zetten rondom de uitgangspunten 'proportionaliteit' en 'hergebruik van data'. Als basis dient hiervoor de Hub-Spoke data architectuur van Gartner. De datareis van NS zal via de oorspronkelijke data architectuur, lessons learned en de opzet van de nieuwe data architectuur worden toegelicht. |

UPGRADE JOUW KENNIS OVER DATA MANAGEMENT

Volgende sessies van de 'Community Data management'

workshop Analytics Translation

Datum: woensdag 29 mei **Tijd:** 17:00 t/m 18:00 uur: inloop & diner 18:00 t/m 21:30 uur: workshop

Locatie: Kantoor Zoetermeer

Wil je deelnemen? Meld je aan per mail bij Raymond Hoogendoorn (raymond.hoogendoorn@the-future-group.com)

| | | | |
|---------------------|---|--|--|
| sessie 4 Q3 2024 | Hoe Business, IT en Datamanagement er samen wel komen | <i>Ronald Baan Dama-NL</i> | De business wil dingen, IT moet het mogelijk maken en Datamanagement moet zorgen dat data blijft stromen en meer dingen mogelijk maakt. Hoe vaak komt het voor dat deze 3 partijen elkaar toch niet begrijpen. Ze spreken andere talen, maar moeten tot elkaar komen om echt vooruit te komen. Ronald schets het moderne datalandschap, bestrijkt het hele gebied van datamanagement, wie wat doet en hoe datamanagement de verbindende schakel tussen de business en IT kan zijn. |
| | Meta data management; De informatie maakt de muziek | <i>Willem Krijgsman / Marco Veerkamp</i> | Maatregelen moeten zorgen voor de bescherming van de bedrijfsgegevens (kroonjuwelen). Te zware maatregelen kunnen impact hebben op performance en kosten, te lichte maatregelen kunnen risico's openbaren voor beschikbaarheid, juistheid, en vertrouwelijkheid. Bij het vinden van de juiste balans speelt gegevensclassificatie een belangrijke rol. |
| sessie 5 Q4 2024 | Model gedreven ontwikkelen | | |
| | Meta data management; Cosmo-girl assesment? | | |
| sessie 6 Q1 2025 | | | |
| | | | |



De evolutie van Data & Analytics oplossingen

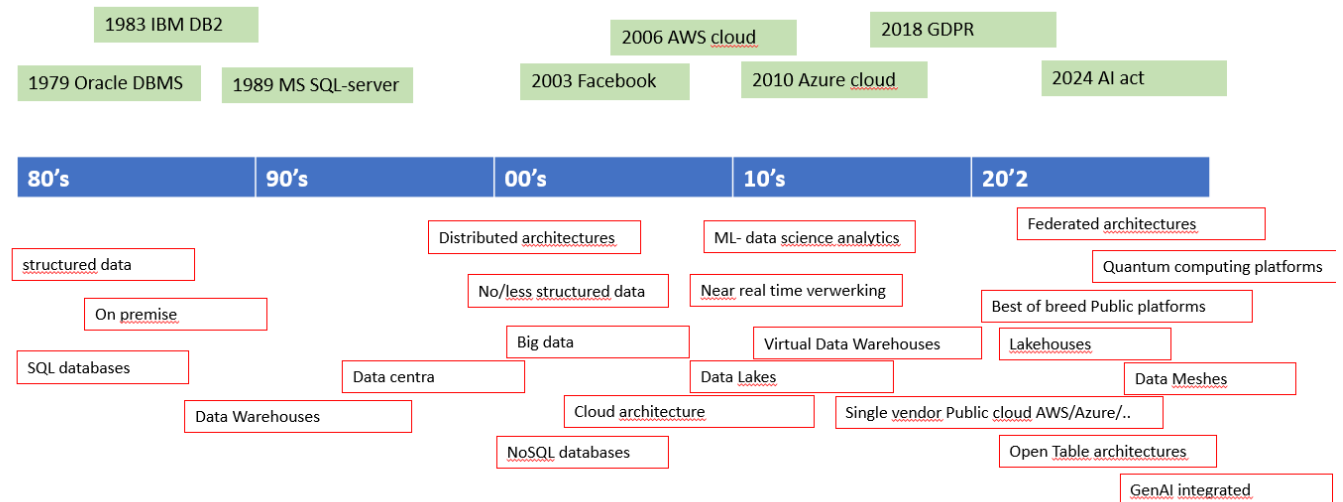
Van Data Warehouse naar Data Lake House en verder

Kennissessie Data Management
14 mei 2023
Eric Fickel

Agenda



1. 1980 – 2000; Business Intelligence & Data warehouses
2. 2000 -2010; Big data en NoSQL
3. 2010 - 2020; Cloud computing en geavanceerde analyses
4. 2020 - verder; Governance
5. Reflecties



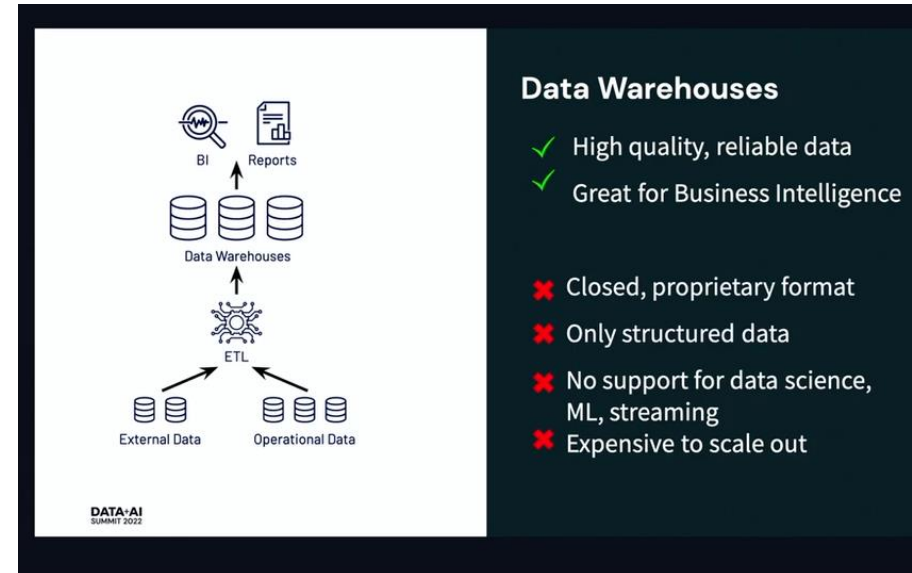
80's – 90's Data warehouses architectuur ter ondersteuning van Business Intelligence



- Gegevens elektronisch opslaan op gestructureerde wijze
- Centraliseren van gegevens opslag in SQL Db
- Rapportages over bedrijfsprocessen
- Verantwoorden, maar ook analyse en sturing
- Sneller en meer ... door gespecialiseerde afdelingen

- SQL databases ; IBM, Oracle en later MS
- Database management systemen (ACID)

- Bill Inmon; één waarheid met centraal DWH; bron gedreven
- Kimball; Data marts in sterren, vraag gedreven
- Gelaagde architectuur; staging, historiseren en business layer
- Dan Linstedt; Data Vault; schaalbaarheid en automatiseerbaarheid

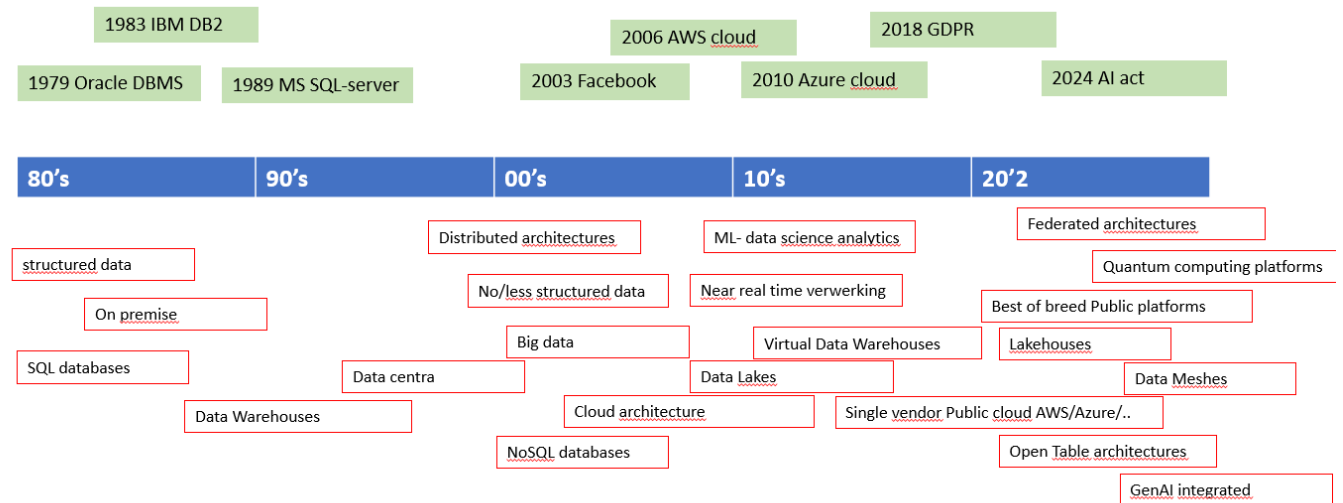


- Minder belasting van operationele systemen
- Historie opbouw
- Centraal beeld van de data van de organisatie
- Eenduidig beeld voor BI om op aan te sluiten
- Back-end (Data store & ETL) en Front-end (verschillende vendor oplossingen BO, Cognos, SAS, etc,) los gekoppeld
- Export naar Excel veel voorkomende functie ☹️

Agenda



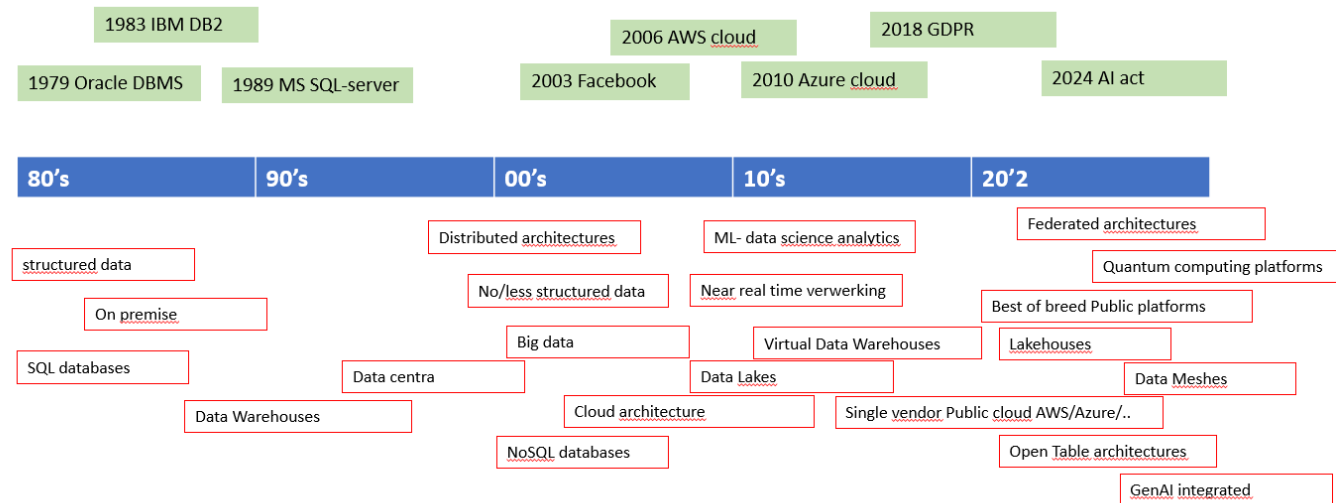
1. 1980 - 2000; Business Intelligence & Data warehouses
2. 2000 -2010; Big data en NoSQL
3. 2010 - 2020; Cloud computing en geavanceerde analyses
4. 2020 - verder; Governance
5. Reflecties



Agenda



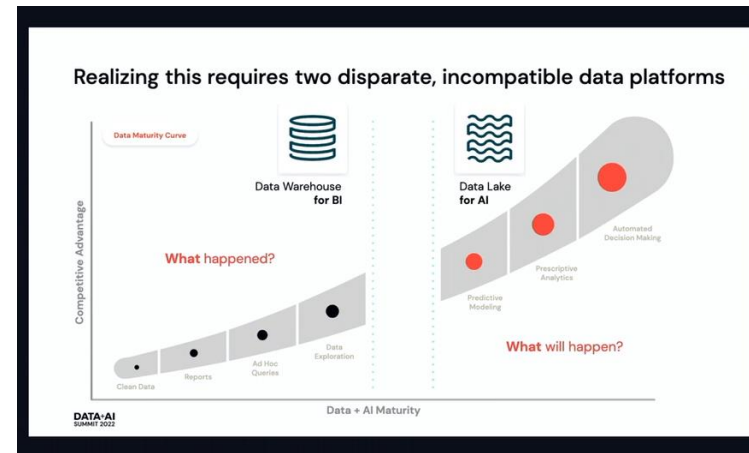
1. 1980 – 2000; Business Intelligence & Data warehouses
2. 2000-2010; Big data en NoSQL
3. 2010- 2020; Cloud computing en geavanceerde analyses
4. 2020 – verder; Governance
5. Reflecties



2010- 2020; Cloud computing en geavanceerde analyses



- Business wil verder ; Voorspellen & Advanced Analytics methoden
- SQL en NOSQL gegevens kunnen combineren en gebruiken
- Flexibilisering en scalability infrastructuur; Opslag, verwerking en analyse van allerlei type data
- Virtualisatie van DWH (onderdelen)
- Microsoft introduceert PowerBI als webtoepassing
- Open source toepassingen o.a. Apache/Spark
- Cloud based DWH platforms o.a. Snowflake
- Graph database o.a. Neo4j, Titan(faya Stax) opkomend populair ivm AI toepassingen
- Big Tech ontwikkelen cloud technologie in eerste instantie voor zichzelf maar bieden vervolgens ook deze technologie aan; Google Cloud, Amazone Web services, MS Azure
- Cloud maakt het mogelijk voor bedrijven om op te schalen en nieuwe technologie in te zetten zonder grote 'up front' investeringen

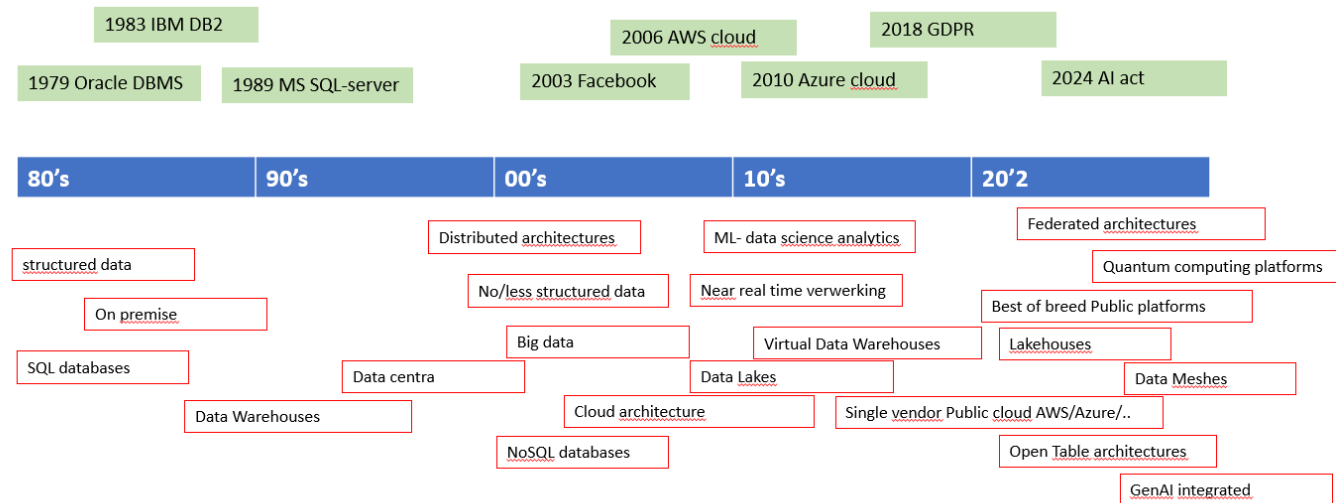


- **Grote gegevensbronnen:** uitdagingen op het gebied van gegevensopslag, -verwerking en -analyse. Big data-technologieën zoals Hadoop, NoSQL-databases en cloud computing spelen een cruciale rol bij het mogelijk maken van de verwerking van grote volumes gegevens.
- **Machine learning en kunstmatige intelligentie:** Data science gebruikt geavanceerde machine learning- en AI-algoritmen om voorspellende analyses, patroonherkenning, clustering en classificatie uit te voeren op gegevens.
- **Data science tooling:** Een opvallende ontwikkeling was de notebooks, interactieve omgevingen waarin code, visualisaties en tekst naast elkaar kunnen worden gebruikt. O.a. Jupyter Notebook.
- **Open source gemeenschap:** De data science-gemeenschap, samenwerking en kennisdeling. Populaire bibliotheken zoals Python's pandas, scikit-learn, TensorFlow en PyTorch

Agenda



1. 1980 – 2000; Business Intelligence & Data warehouses
2. 2000-2010; Big data en NoSQL
3. 2010- 2020; Cloud computing en geavanceerde analyses
4. 2020 – verder; Data platforms en Governance
5. Reflecties



2020 – verder; Data platforms en Governance

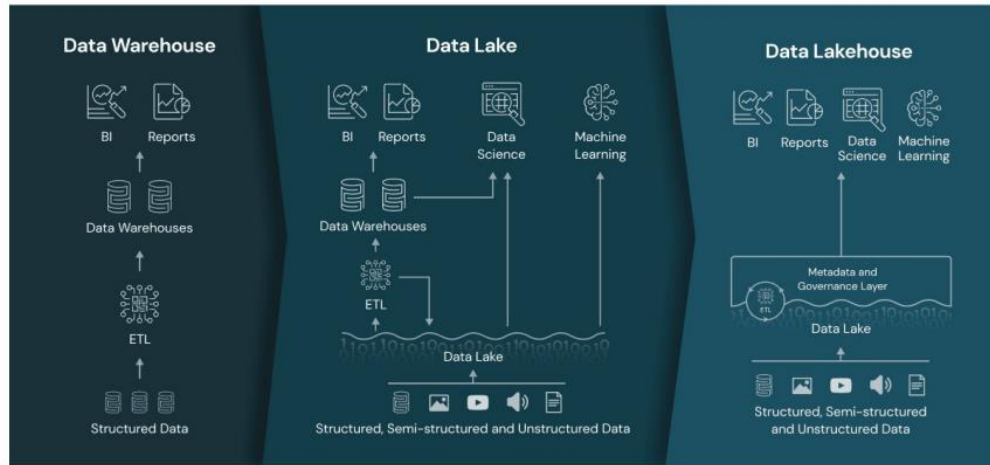


- Data als een product; Data Market place, fabrics etc. Business heeft als doel data te laten gemakkelijker en sneller beschikbaar te stellen aan verschillende doelgroepen
- 2018 GDPR zorgt voor veel aandacht governance aandacht in de 20's
- Citizens data science; business krijgt mogelijkheden van advance toepassingen o.a. GenAI.

- **Data Lakehouse, Data Mesh, Data Products/ Data Fabrics**
- Single vendor vs Best of Breed stacks
 - Multiple approaches to create data as a product; one storage, multiple zone/domains, zoned lake houses, etc.
 - Open Table standards => multiple query engines/vendors
- Centralized vs federated
- DataOps/ CDO/ Data organisatie
- Use of GenAI/ LLMs for processen & analyse un/semi structured data

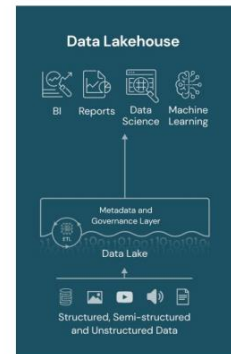
De data lakehouse; best of two worlds?

Het DAP'R plateau 1.0

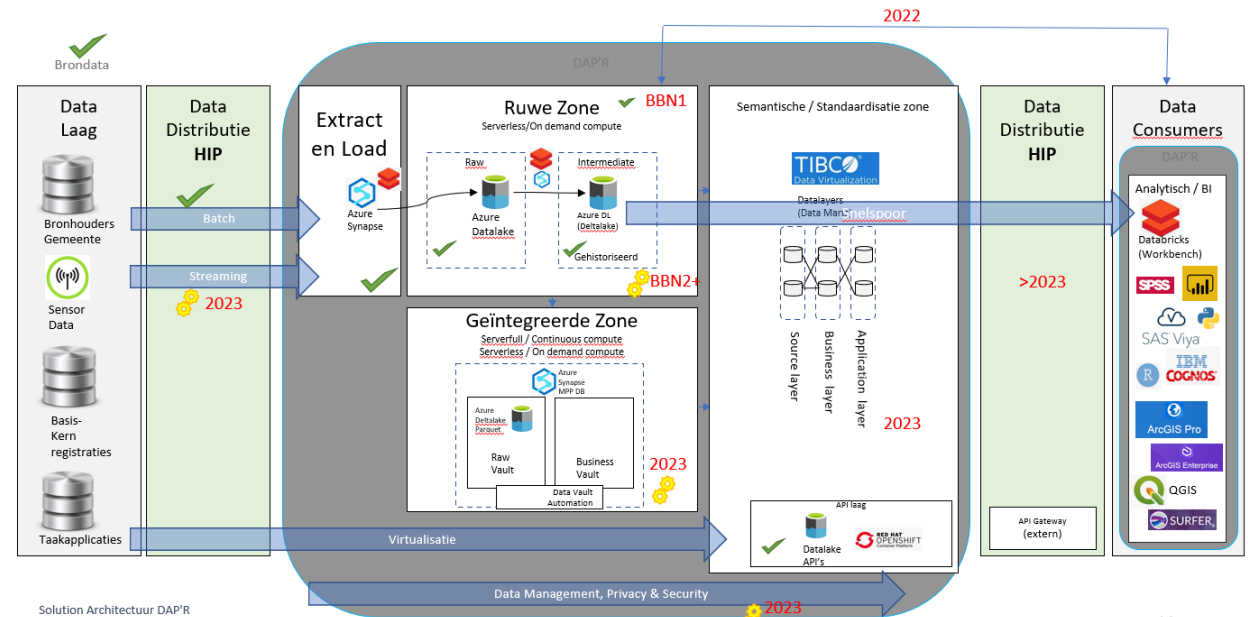


Source: Databricks.com

Key Characteristics of a Data Lakehouse



- Two use cases: BI and data science
- Data is stored once
- Supports structured and unstructured data
- Schema enforcement
- Open file formats
- Low-cost data storage
- ACID compliant



Solution Architectuur DAP'R



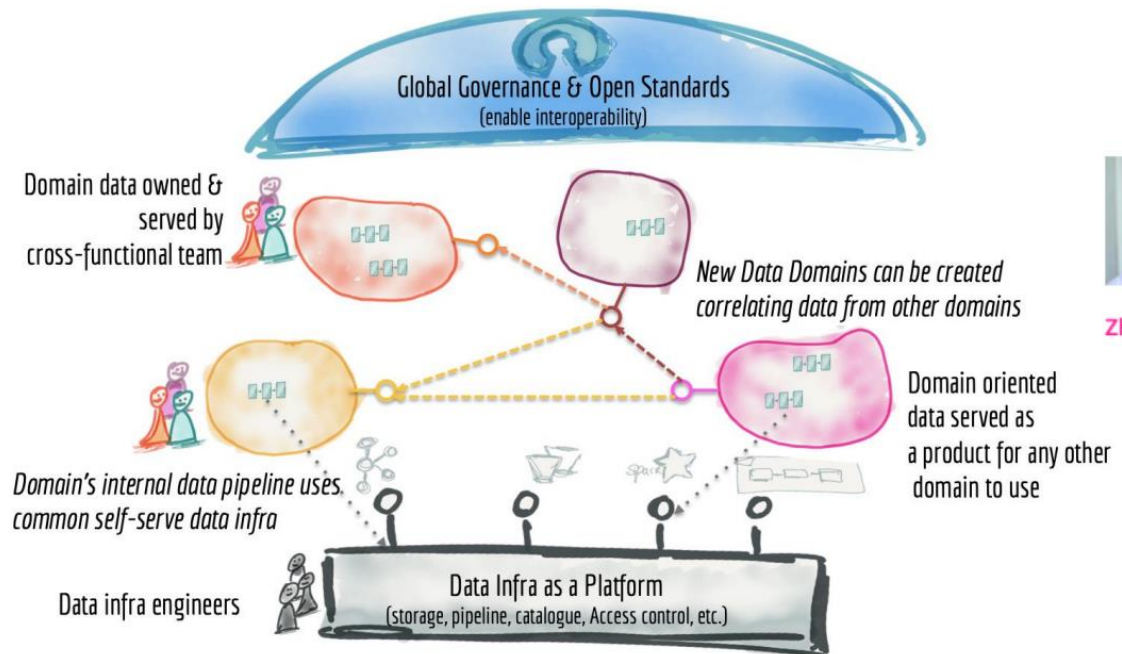
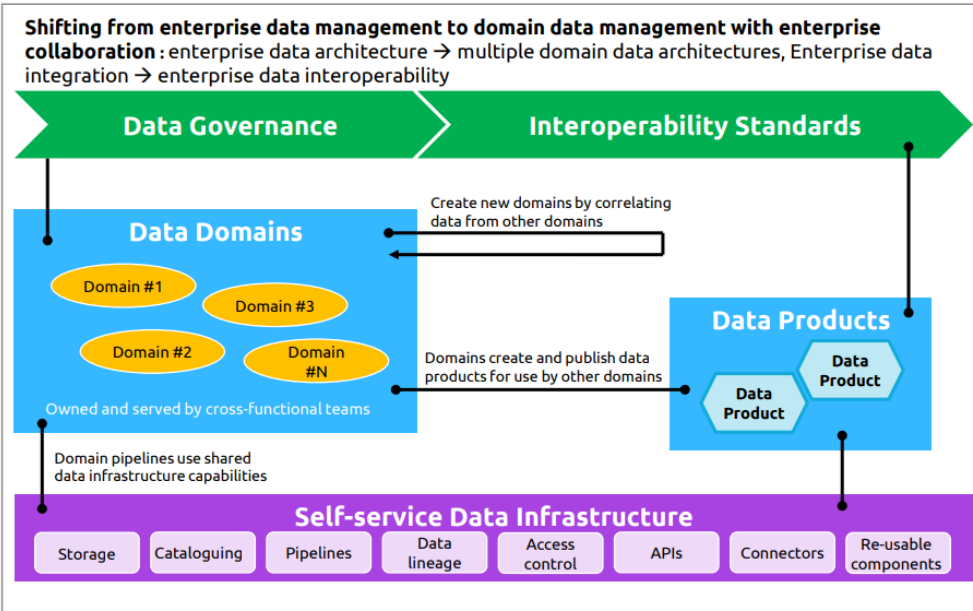
Delta Lake allows you to *incrementally* improve the quality of your data until it is **ready for consumption**.

Data Mesh

Data Mesh verwijst naar het concept van het federatie van eigendom van gegevens naar domein-georiënteerde teams en het gebruik van een gestandaardiseerde set blauwdrukken voor self-service, interoperabiliteit en multi-vorm gegevensopslag en consumptie.



DATA MESH PRINCIPLES



Zhamak Dehghani

Data products; data beschikbaar als een product voor inter- en externe afnemers (Mike Ferguson)



Steps To Creating For Data Products
 – From Data Concept Model To Data Marketplace

1. Identify the data concepts, properties and relationships and construct a data concept model
2. The data concepts become the 'skeleton data entities' in the common vocabulary
3. Each data concept and its attributes should be defined in the business glossary as a data entity with a data owner
4. Use the catalog to discover the data for each data entity in underlying data stores across the data landscape
5. Design DataOps component-based pipelines to create the data products with common vocabulary data names
6. Publish all data products in a data marketplace

Data Concept Model Example

Copyright © Intelligent Business Strategies 1992-2024

A Data Foundation Includes Reusable 'Business Ready' Data Products That Are Semantically Linked, Built Once And Reused Everywhere For Different Analytical Workloads

Deeper integration is happening between DWs and data lakes that can also reduce data redundancy

Online behaviour analysis
 Customer sentiment analysis
 Customer segmentation analysis
 Fraud analysis
 Claims and payments, Marketing analysis

Copyright © Intelligent Business Strategies 1992-2024



Top Ten Key Trends In Data Management And Analytics (D&A)



1. Generative AI – innovation in every area of business
2. Hybrid multi-cloud computing is now the norm
3. Architecture modernisation - integration of data warehouses, lakehouses, data lakes & streaming
4. Rationalisation towards a common data & analytics software stack for development & governance
 - Do more with less, integration across tools, accelerated development and shared metadata
5. FinOps - CFOs demand visibility of the full cost of the D&A ecosystem with consumption-based pricing
6. Data governance remains very high priority with AI governance now also on the agenda
 - More sources, distributed data complexity, data quality, security, privacy, usage, observability, sharing, retention
 - Poor data culture and weak data governance are barriers to decentralised development
7. Increasing demand for lower and lower latency data
8. Democratisation and acceleration of data and analytics development
 - E.g., Data Mesh, citizen data engineering, DataOps, autoML, MLOps, CI/CD, code Vs low code / no code
9. Compliant sharing and reuse of data and analytical products in data marketplaces / exchanges
10. Growth in intelligent applications, decision intelligence and AI-Automation

Data governance as keystone for compliant AI and digital trust

Jan Henderyckx



- “Monkey with the hand grenade” moment
 - Incidenten met GenAI exponentieel toegenomen
 - EU AI Act vs US approach to AI
- Handling the Monkey
 - AI Live cycle approach
 - Mature the AI governance framework
 - FAT compliant (Fairness, Accountability, Transparency)
 - Apply critical thinking
 - Data quality; what relevant dimensions to govern



Reflecties

- Data democratisering is een dringende behoefte van de business;
- Beheersen en beheren van data en het gebruik is een steeds grotere taak/zorg;
- Data architecturen zijn er in alle soorten en maten;
- Kies niet wat er is, maar kijk goed naar volwassenheid organisatie waar je het voor opzet;
- Kijk uit voor vendor lock-in. kies voor zo open mogelijk platform;
- Een goede data exploitatie is belangrijker dan altijd maar mee gaan met nieuwe trends;
- GenAI: nieuw turning point?

Een overzicht globaal in de tijd

