Building Sovereignty
A Two-Year Journey to a Scalable Data & AI Platform

2026, March

Antoine Stelma – Co-founder @ Connected Data Group

WE LIVE DATA

CONNECTED DATA GROUP

AGENDA

- Introduction
- The Vision & The Why
- The Infrastructure Challenges
- The Blueprint & Solution Components
- Lessons Learned & Conclusion

ANTOINE STELMA

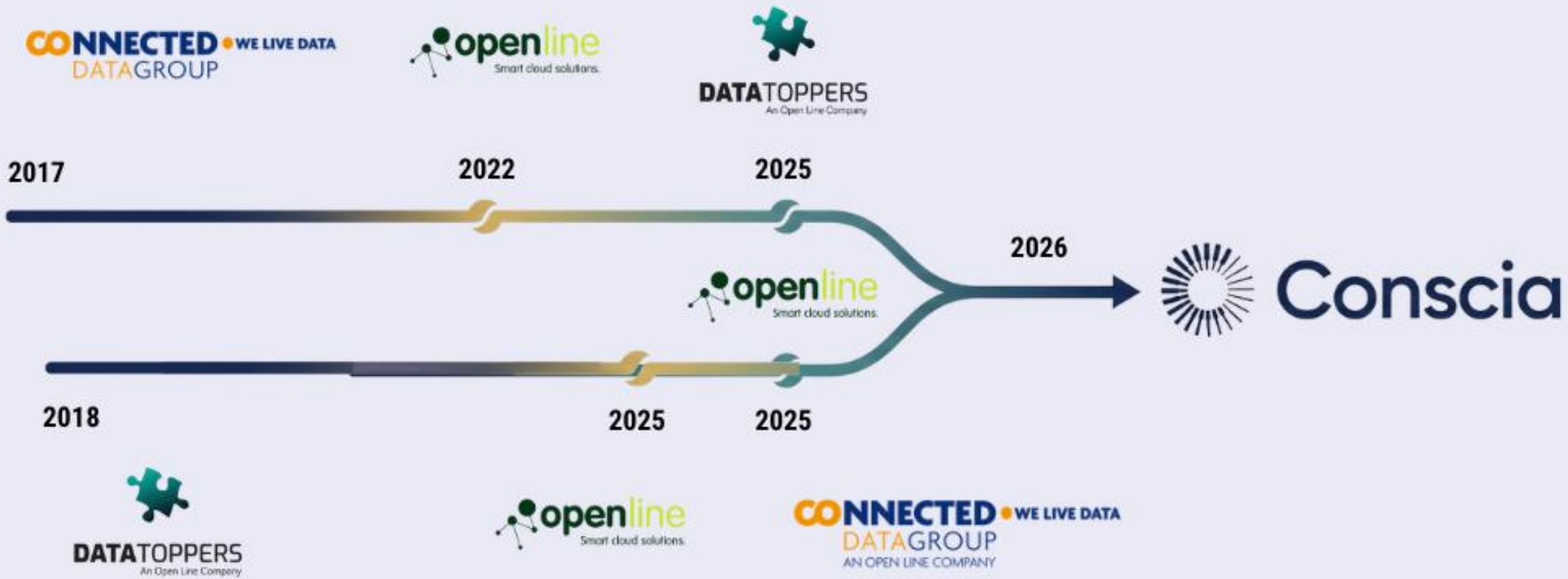
- **20+ Years of Data Mastery:** A seasoned Data Architect and trainer with over two decades of experience in Data Warehousing and Business Intelligence.

- **Co-Founder & Visionary:** Built Connected Data Group alongside Erik Fransen to provide organizations with smarter, sovereign Data & Analytic architectures.

- **Specialist in Scalability:** Deep expertise in Data Vault Modeling, Data Virtualization, and the blueprinting of modern, high-performance data platforms.

- **Global Educator:** Translating complex theory into practice as an instructor for the Connected Data Academy.

Real-World Grounding: Passionate about skipping the "vendor fluff" to share best practices rooted in extensive research and daily, field-tested use cases.


CONNECTED WE LIVE DATA
DATAGROUP
openline
Smart cloud solutions
DATATOPPERS
An Open Line Company
2017
2022
2025
2026
openline
Smart cloud solutions
Conscia
2018
2025
2025
DATATOPPERS
An Open Line Company
openline
Smart cloud solutions
CONNECTED WE LIVE DATA
DATAGROUP
AN OPEN LINE COMPANY

2026 → Conscia

CONNECTED DATA GROUP
WE LIVE DATA

CONNECTED DATA CLOUD
WE EMPOWER DATA

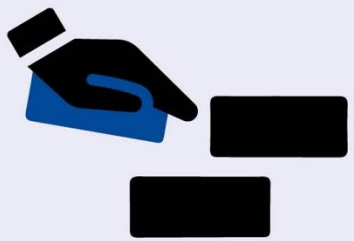
CONNECTED DATA ACADEMY
WE TEACH DATA

WHAT IS OUR RECIPE

TOGETHER WE CREATE
VALUE FROM DATA

GRIP

ORGANIZING DATA



ANALYTICS PLATFORM

SOLUTION AS A SERVICE

INSIGHTS

ADDING VALUE

KNOWLEDGE

CONNECTED DATA ACADEMY

The Vision & The Why

The Ambitious Question

Why should we build a truly scalable sovereign platform?

- Escape from Vendor/Cloud lock-in
- Compliance as a Competitive Advantage
- Standardize with best-of-breed solutions
- Control of cost
- AI readiness & Intellectual Property Protection
- Agility through Hybrid Flexibility

A definition of sovereignty

True Sovereignty is the **strategic capability** of an organization to maintain **absolute authority and self-determination** over its **digital landscape** while navigating complex regulatory and technical environments.

Three core ownership pillars:

- Ownership of Logic
- Ownership of Data
- Ownership of Solutions

The Market Gap

Why existing off-the-shelf solutions weren't enough.

- The "Black Box" Logic Trap
- Mandatory Vendor Lock-in (The "Cloud Hotel California")
  - Most modern SaaS data platforms are "hotel" architectures:

     ‘You can check out every time you like. But you can never leave’
- The Trap of Unpredictable Cost

The vision

“Create a scalable and flexible Data & AI platform that is Cloud Agnostic, based on best of breed solutions and provides full control over data, logic and solutions.”

“In our vision, data, logic and AI models are transparent and move with the business, not the vendor.”

“Data Governance & Data Management are continuing processes and not technical solutions. How can we support the process?”

The Infrastructure Challenges

Challenge 1: Scalability

Kubernetes: the backbone of the operation.

(non-technical explaination)

- Each application is a container
- Data & AI Platform is the cargo
- Each customer has its own private and secured dock
- Kubernetes is the harbour master
- Customer owns the containers and the private dock

CONNECTED DATA GROUP

Challenge 2: Security

Moving beyond the perimeter
—security in a sovereign environment.

(Just some of the features)

- Private Endpoint Connectivity only
- Context-Aware Access Rules per customer
- Advanced Kubernetes Protection
- Security Matrix for users and usage
- Our Security Operation Centre

Security by design is NOT only the platform but starts with the governance of the data, logic and applications

CONNECTED DATA GROUP

Challenge 3: Governance & Compliance

How to bake Governance & Compliance into the platform's DNA.

• Automated Data Lineage	- Track every data movement, 100% transparant
• Metadata-Driven Guardrails	- Using a Datacatalog to enforce policies
• Immutable audit trails	- Log every action taken by users and processes
• Standardized Data modelling	- Data Vault as our core standard
• Proactive Risk Monitoring	- Security Operation Centre
• Lifecycle Management	- How long may we use the data?
• Purpose Limitation	- When it's allowed to use this data?
• Data Minimisation	- Do we realy need this data here also?

Challenge 4: The Cloud Dilemma

Is there a Cloud Dilemma or should we navigate better?

- Our Data & AI platform is entirely cloud agnostic.
- Public Cloud for Scalability is an option
- Private Cloud for sovereign data for sensitive data and core business logic
- Hybrid as best of both worlds is always possible

But remember the Cloud hotel California:
It’s only a good hotel if you check out and really leave.

Challenge 5: The Tooling Balance

The tug-of-war between Open-Source & Closed-Source application

Challenge 5: The Tooling Balance

- **Closed Source**
  - Do we need support from the supplier?
  - Exit scenario always results in readable code, logic and data for fast migrations
  - Operational Efficiency versus the cost?

- **Open Source**
  - Can we support the application?
  - Exit scenario always results in readable code, logic and data for fast migrations
  - Is there an active community that we can support?

Challenge 6: Delivery Models

Software as a Service versus Product as a Service

SAAS
Customer uses the Data and AI platform as a service.

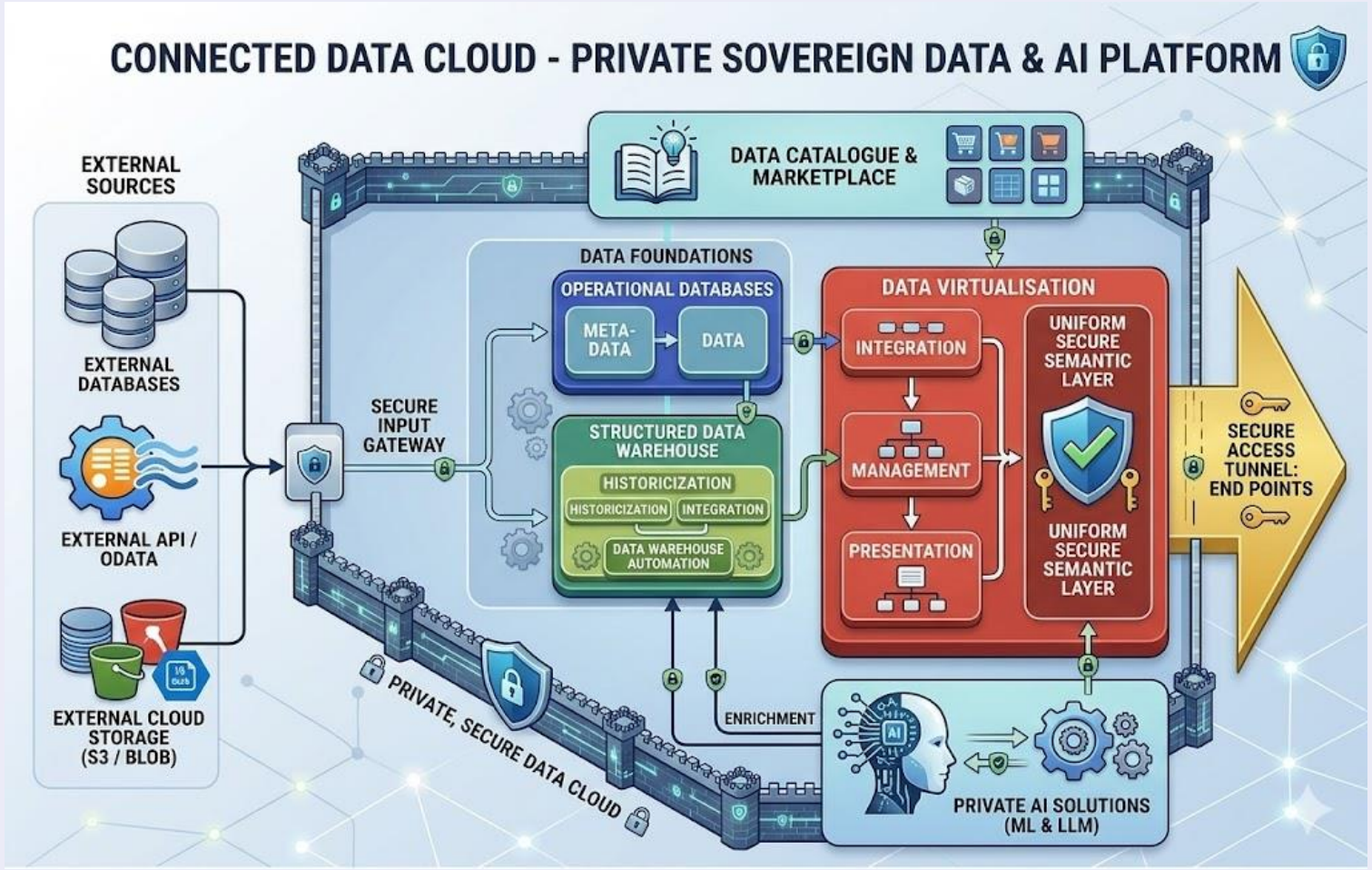
PAAS
More customers are now using the dataproducts that are generated with our platform.

We provide housing corporations with **VERA-based sovereign data warehouses** as a fully managed service. This includes 150 KPIs and 7 CORA reports to satisfy **mandatory government reporting**, alongside daily data loading, proactive maintenance, and regular updates.

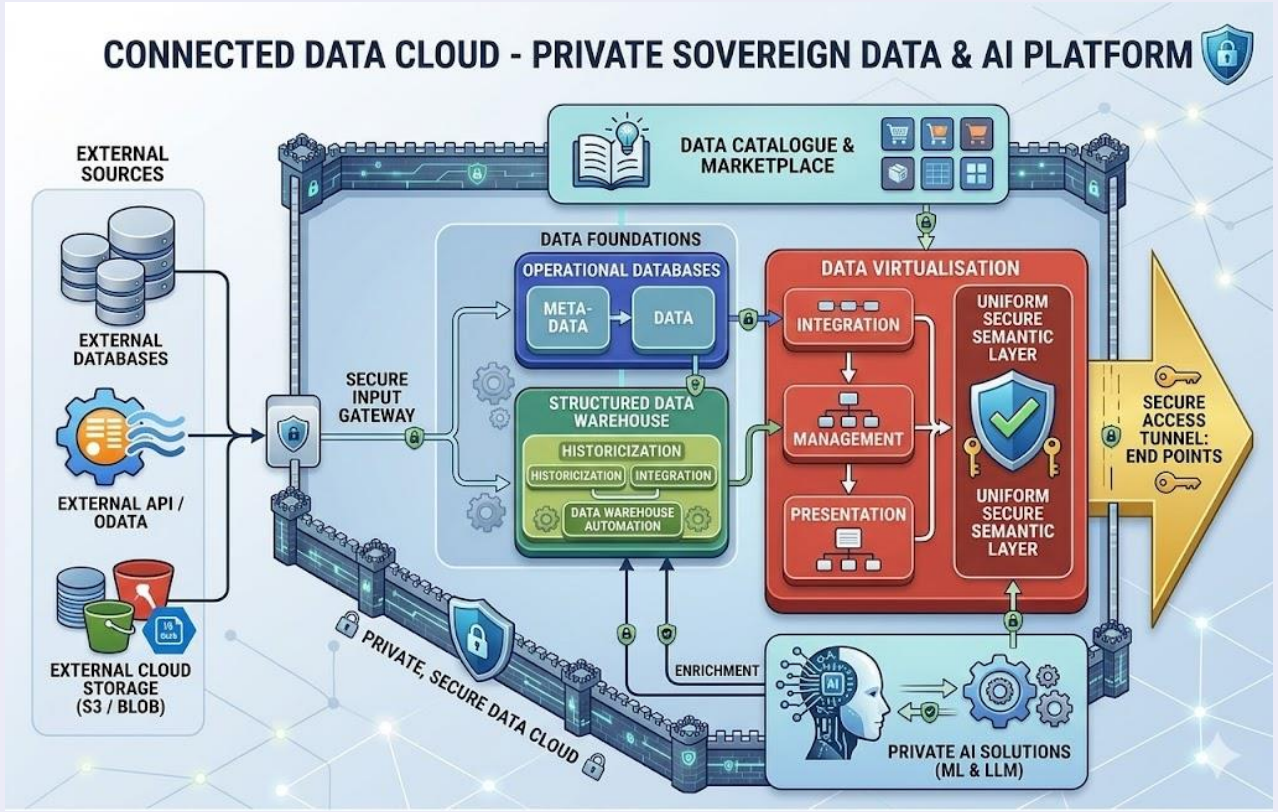
Both models use the same data and ai platform

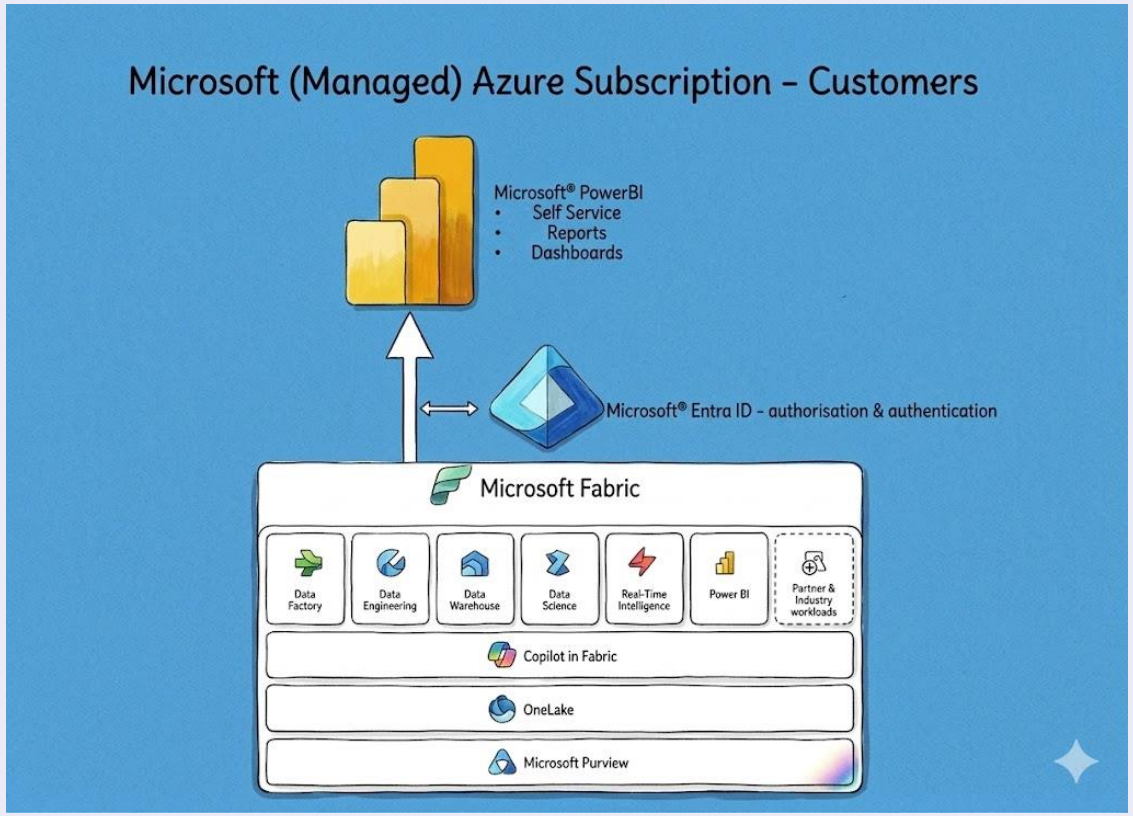
The Blueprint & Solution Components

The Big Picture - Sovereign



The Big Picture - Hybrid





An architectural diagram of the resulting platform hybride with MS Fabric.

Component 1: Data Warehouse Automation

Our demands

• Automation Excellence	- Proven track records in automating Data Vaults
• Native Elasticity	- Able to run as a scalable container
• Continuous Quality	- Delivery of code and data products
• Progressive Ecosystem	- Active supplier of community to support our demands
• Performance Driven ROI	- License cost should match the business case for saving
• Deep Interoperability	- Integrate with any source
• Standardized Modelling	- Standardization to enhance auditability and historical integrity
• Commercial Safety	- For closed-source a comprehensive SAAS license must be available

All products create with the automation solution must be 100% transparent.

Component 1: Data Warehouse Automation

Our choice

With an impressive background in Swiss banking and a strong Data Vault track record, we choose Datavault Builder as our strategic partner.

Component 2: The Data Catalog

Our demands

- Align with enterprise Data Governance standards
- Automate the ingestion of metadata
- Enable data quality observability
- Provide comprehensive support for data domains
- Maintain a centralized metadata glossary
- Implement automated data tagging by subject area
- Support fine-grained, role-based access control
- Streamline workflows for resolving data quality issues

Component 2: The Data Catalog

Our choice

Open Metadata (Open Source)

- Active Community Support
- Comprehensive Data Management
  - Seamlessly handle Discovery, Lineage, Observability, Quality, Collaboration, and Governance.
- Centralized Metadata
  - All data is stored in a Postgres database, ready for Reporting and AI integration.

Component 3: Data Virtualization

The "Magic Trick"—accessing data without moving it.

Our demands:

- Zero-Movement Access: Enable real-time data access without costly migrations
- Minimized Redundancy: Eliminate data duplication to streamline storage
- Unified Logical View: Speed up exploration and time-to-market via a single interface
- Virtual Agility: Accelerate development using flexible virtual data models
- Direct Governance: Centralize security by applying policies directly at the data source
- AI Acceleration: Securely fast-track AI use cases through unified data exploration

Component 3: Data Virtualization

Our choice

Denodo Data Virtualization

- Dedicated supplier
- Fully audible event logs
- Data Marketplace for direct integration with generative AI
- Proven track records in the Virtual Data Management market
- Transparent backend with readable code for exit scenario’s
- Native scaling via container technology
- Excellent support in the Netherlands including a Dutch User Group

Component 4: Databases and the Data Lake

The foundation of the storage layer.

Our demands:

- Standardize on PostgreSQL for all organizational database management
- Utilize JSONB for high-performance object storage in a relational model
- Leverage native geospatial support for location-based analysis
- Ensure ACID compliance for data integrity across critical workloads
- Optimize performance through advanced indexing and execution plans
- Secure sensitive data with Row-Level Security (RLS) and enterprise controls
- Reliable Metadata Store: Balance high performance with robust backend stability
- Containerize deployments to increase infrastructure agility

Component 4: Databases and the Data Lake

Our choice

Postgres and Single Store

SingleStore: Real-Time Performance

- High-Speed Analytics for real-time analytical requirements
- Hybrid Data Models for SQL and NoSQL functionalities
- In-Memory Processing: Utilizes in-memory capabilities for low-latency operations
- Flexible Storage Architecture: Supports both row-based and column-based storage formats
- Multimodal Capabilities: A single database solution for time series, full-text search, relational, and spatial data

Component 5: AI solutions

Bring the models to the data, not vice versa

Our demands:

- Bring the AI to the Data:
  - Developing private ML, Generative & Agentic applications
- Dutch Sovereignty:
  - All solutions are hosted and secured within our local Dutch datacentres
- Security First:
  - A dedicated Data & AI platform built for safety and customer trust

Component 5: AI solutions

Our journey sofar

- Empowering Research by supporting our customers Data Science team with our solutions
- Specialized Knowledge Systems: Developing a high-performance RAG application for complex document knowledge retrieval
- Multi-Model Versatility: Leveraging the Ollama framework to deploy and manage a diverse range of Large Language Models (LLMs) and Vision Language Models (VLMs)
- Cutting-Edge Integration: Engineering Model Context Protocols to bridge Generative AI with Data Vault architectures
- Continuous Innovation: Cultivating a culture of daily learning and technical growth

Lessons Learned & Conclusion

Lessons Learned

What we got right - The "Aha!" moments that validated the 2-year journey

- **Strategic Partnerships**: Identified sponsors who provided both organizational support and the expertise needed to transition our platform into a managed service
- **Conceptual Alignment**: Assembled a multi-disciplinary expert team to ensure technical capabilities are always aligned with core business functionalities
- **Methodology as a Product**: While technical specs are vital, our greatest success is embedding our methodology and best practices directly into the platform
- **Outcome-Driven Focus**: Prioritizing the "human" side of the platform—methodology and best practices—as the ultimate driver of customer success

Lessons Learned

What We Got Wrong. Honest reflections on the hurdles that slowed the team down

- **The Scalability Gap:** We waited too long to formalize knowledge transfer, which created a bottleneck for our Managed Service transition
- **The Evolution:** Moving from a stand-alone **solution** to a full **ecosystem** proved to be a more complex journey than anticipated
- **The 100% Rule:** While we aimed for perfection, our customers were actually willing to partner with us and grow alongside the product
- **Partnership vs. Product:** Real-world feedback during the "journey" was more valuable than a delayed, perfect launch

Conclusions

3 Key Takeaways

- **Sovereignty is Strategy** True digital sovereignty isn't a buzzword; it’s the strategic capability to maintain **absolute authority** over your digital landscape
- **Flexibility Over "Black Boxes"** Avoid being locked into one cloud. Use a **cloud-agnostic blueprint** that blends Public Cloud scalability with Private Cloud security
- **Methodology Drives Success** Value comes from the platform’s **governance and standardization**, not just the technical specs or the data itself

Question & Answers

CONNECTED

DATA GROUP

WE LIVE DATA