

# Half a million rows later How we modernized our analytics platform in 6 months

Sébastien Stampfli (CIO)

Martin Meister (PO Data Systems and Services)

18. September 2025

# Content

The SNSF and its data

Why we had to change

The planned modernization strategy

What actually happened

What we learned

# Swiss National Science Foundation (SNSF)



promotes excellent scientific research



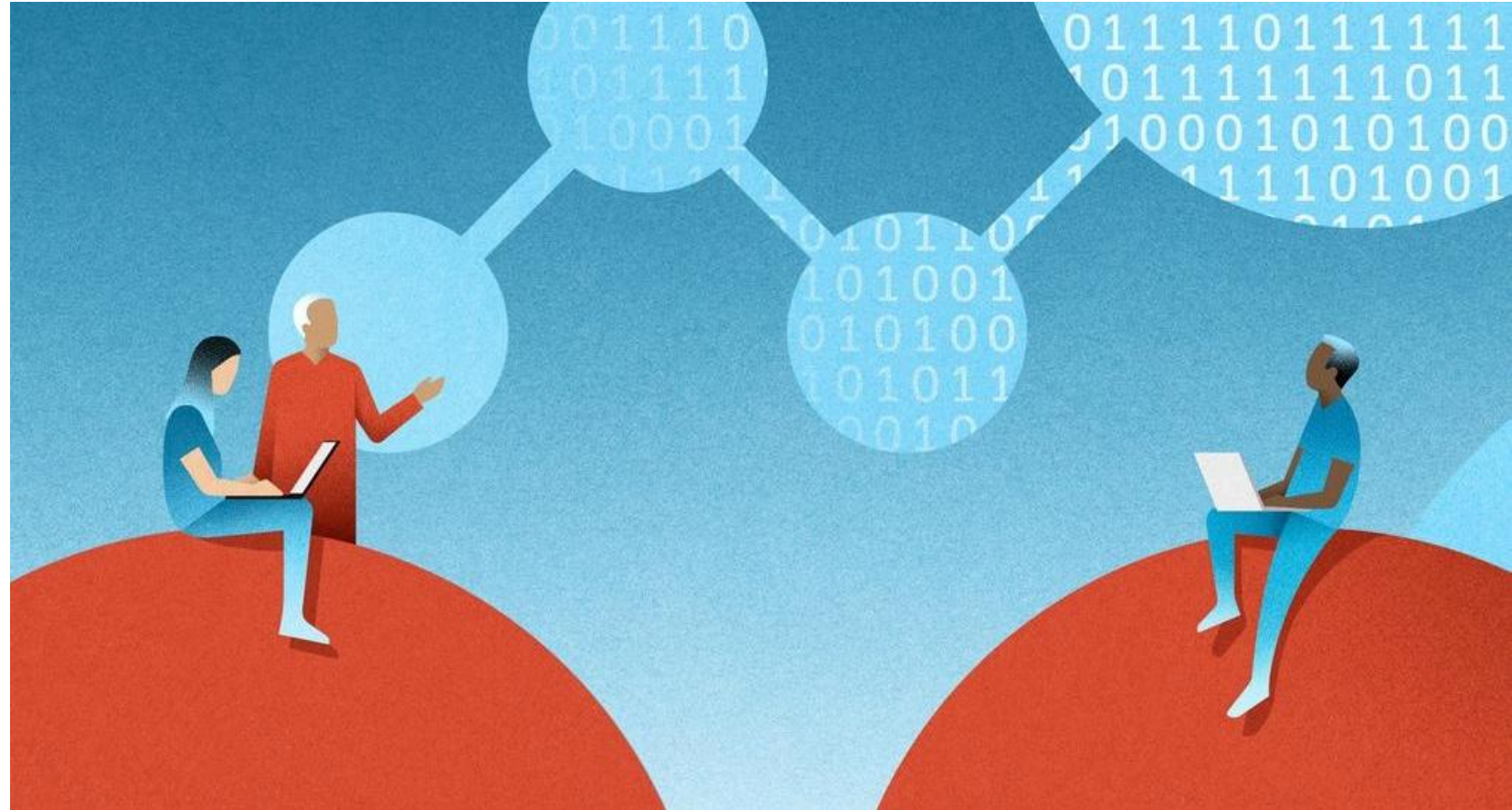
strengthens Switzerland as a centre of science



is committed to sustainable development and equal opportunities

# Data at the Swiss National Science Foundation

“Our data isn’t just technical — it’s a mirror of decades of changing science funding practices.”



# Swiss National Science Foundation's Data Strategy



Augment processes with insights



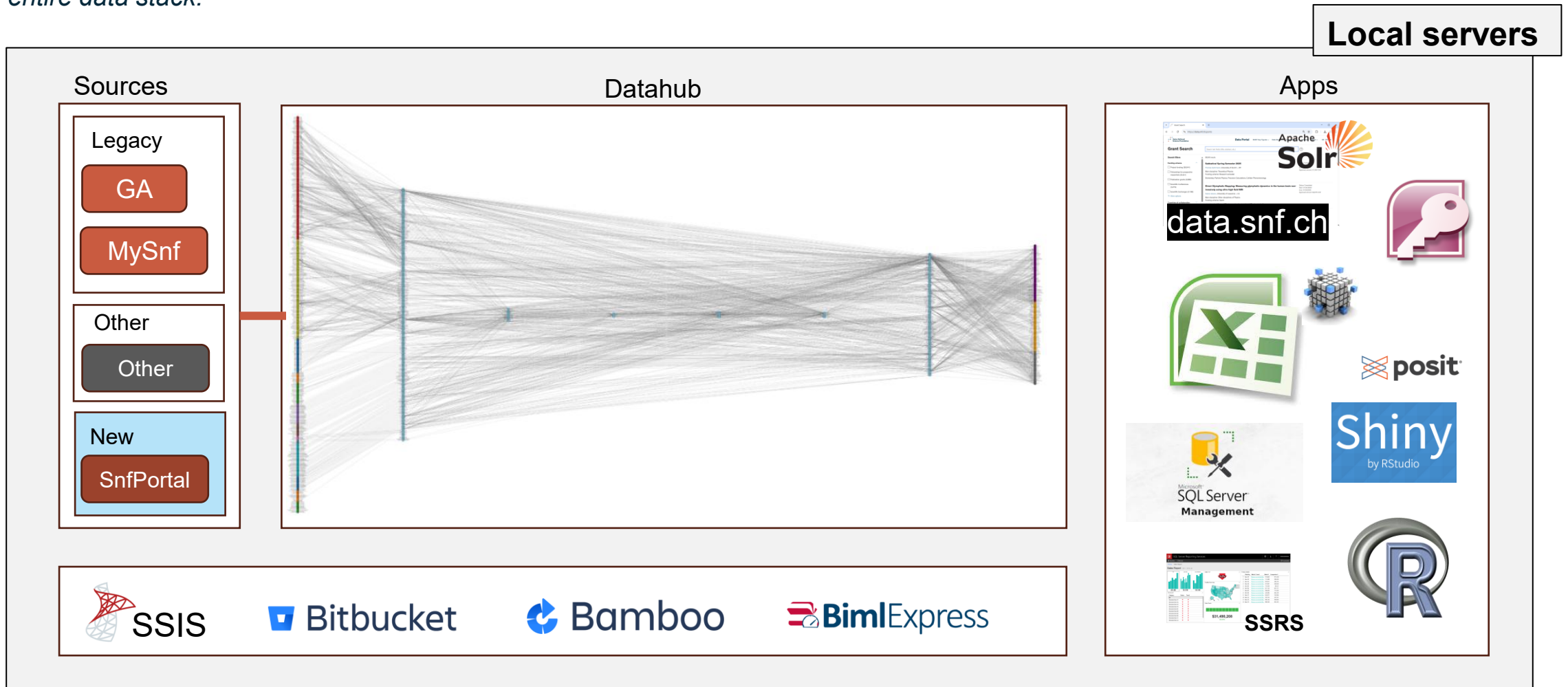
Data-literate organisation



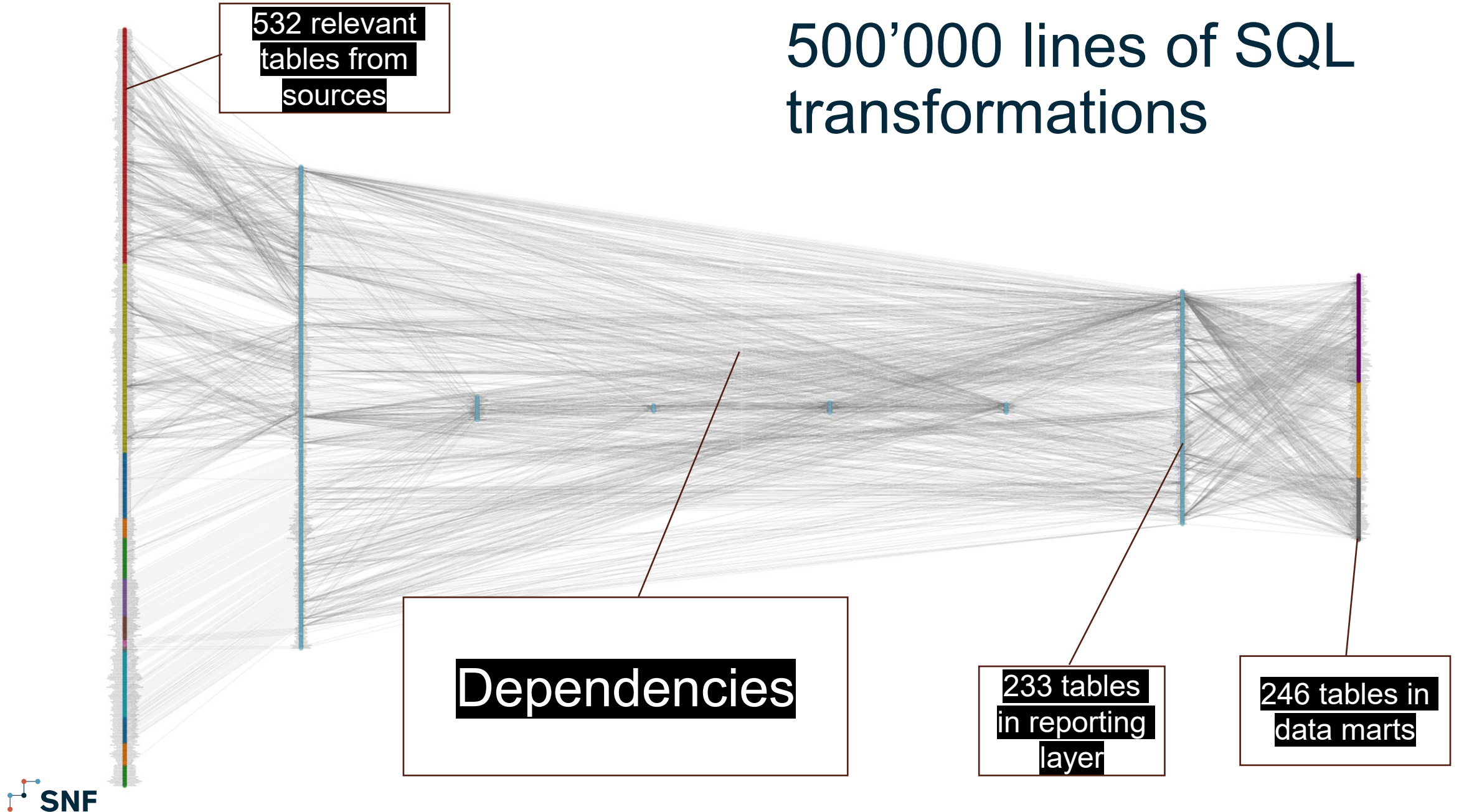
modular and scalable  
self-service data platform

# The old analytics landscape @SNSF

*“Despite having a well-oiled BI machine, major business and technical shifts required us to rethink our entire data stack.”*

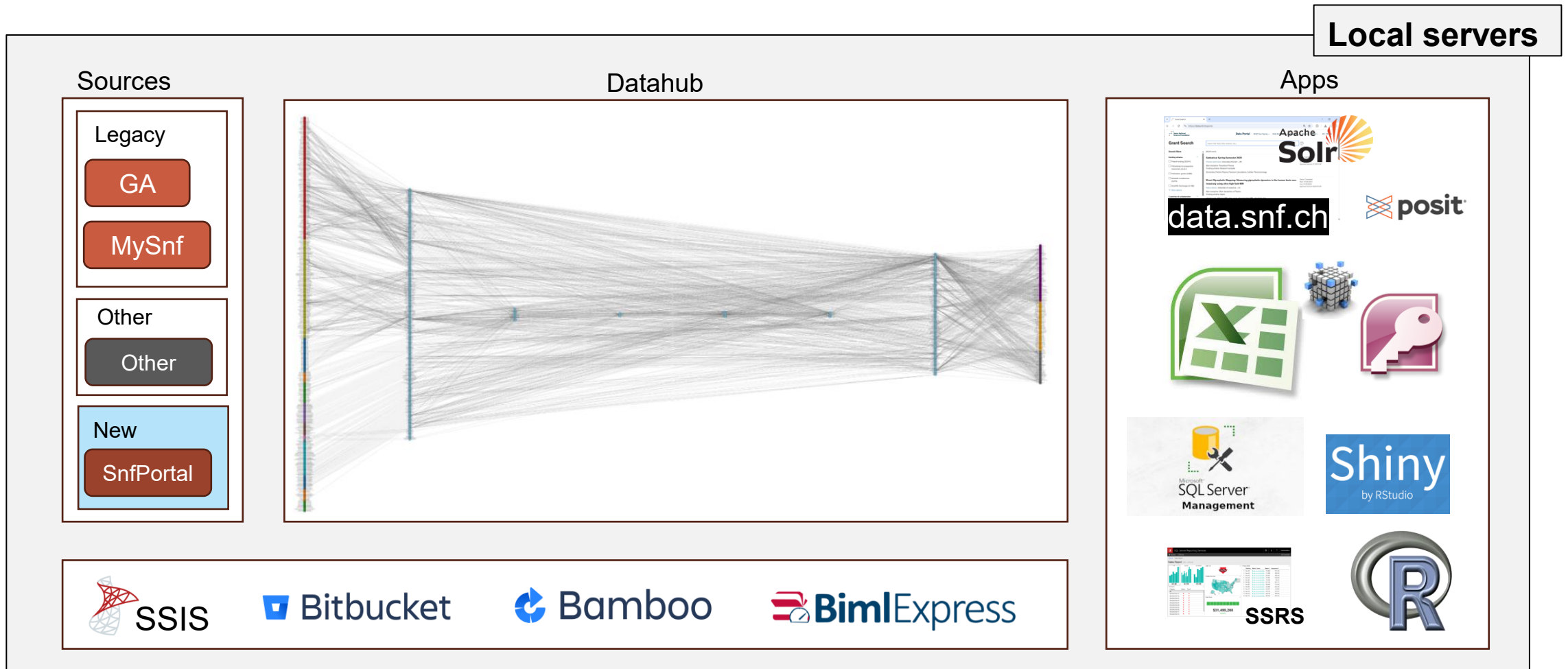


# 500'000 lines of SQL transformations



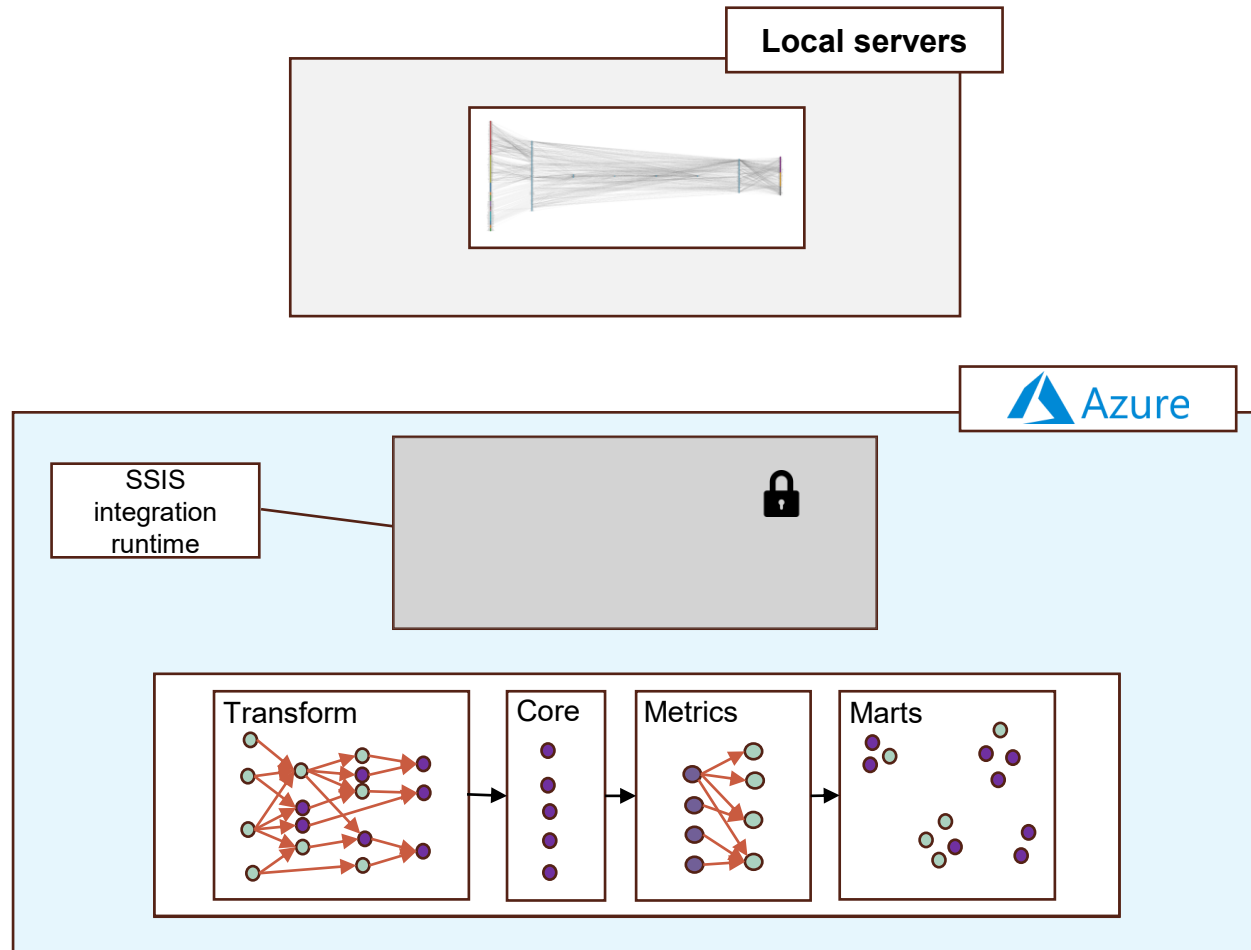
# The old analytics landscape @SNSF

*“Despite having a well-oiled BI machine, major business and technical shifts required us to rethink our entire data stack.”*



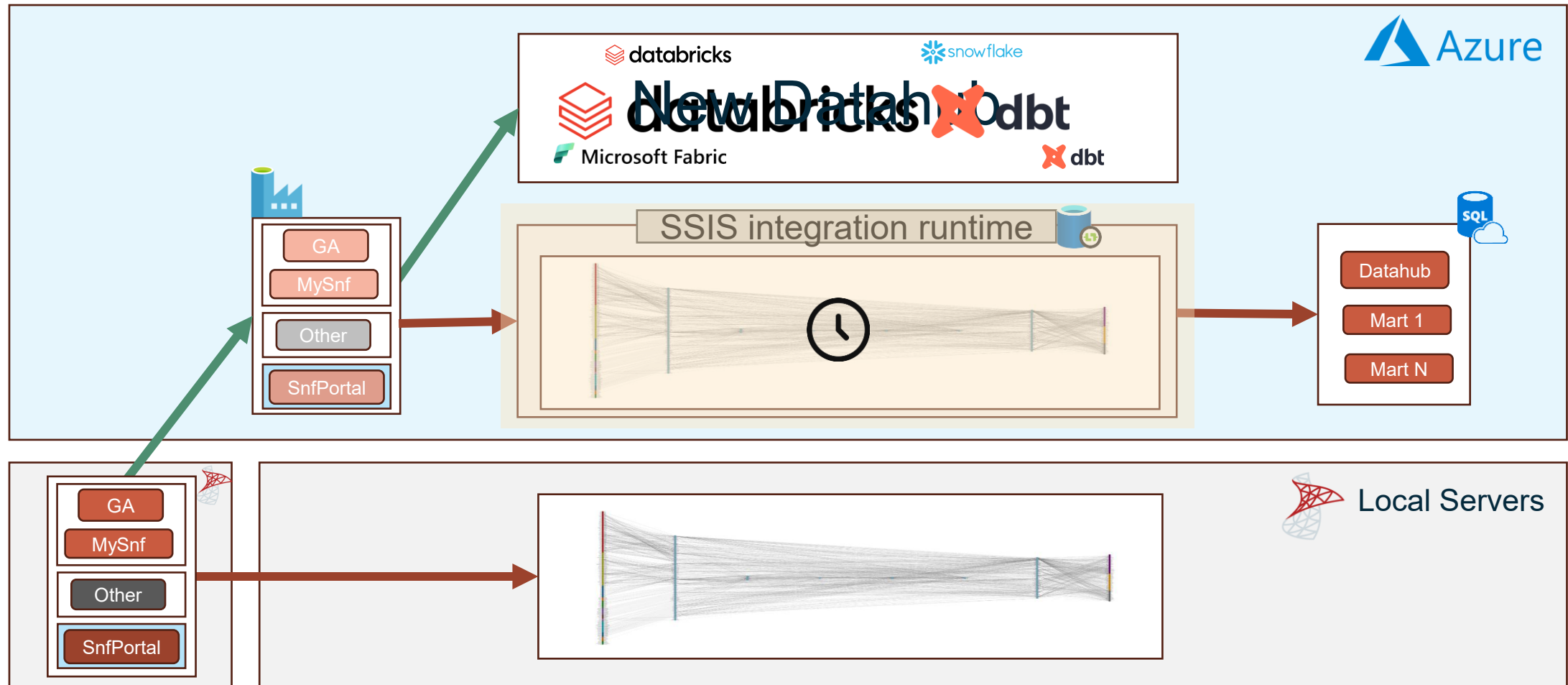
# The planned modernization strategy

*“Our Original Plan was Lift & Shift, Start New & Fade Out”*



# What actually happened

Enabling system platform for SSIS runtime



# What actually happened

Red create is implied

APACHE  
Spark

T-SQL

SSIS int

## From SQL Server to Spark SQL

```
SQLQuery1.sql - sql_datahub-access-ws-prod.database.windows... Quick Launch (Ctrl+Q)
File Edit View Query Project Tools Window Help
master
SQLQuery1.sql - sql...artin.meister (84)
WHEN evaluation.AssessmentType = 'Recommendation' THEN 0
WHEN evaluation.AssessmentType = 'CoRecommendation' THEN 1
ELSE NULL
END
LEFT JOIN [stagesnfportalia].[Invitation] invitation
ON invitation.EvaluationId = evaluation.Id
WHERE evaluation.AssessmentType IN ('Recommendation','CoRecommendation') -- Refe
AND assessment.AssessmentStatus <> 'Canceled' -- Canceled
AND roleMapping.RefereeRoleId = '84D53583-FD67-4308-AE16-5AF0361C6524' /*
AND invitation.InvitationStatus != 'Canceled' /*canceled*/
GROUP BY evaluation.ProposalId
),
stage040_Cv AS (
SELECT
c.*,
CASE
WHEN c.Gender = 'Female' THEN 'f'
WHEN c.Gender = 'Male' THEN 'm'
WHEN c.Gender = 'NonBinary' THEN 'x'
ELSE NULL
END AS GenderAbbr
FROM stage040_Cv c
),
withdrawalBeforeCallEndDate AS (
SELECT
Proposal.Uuid AS ApplicationId,
StateDatahub.StateDatahubId,
StateDatahub.StateDatahub,
Proposal.Entered,
Call.EndDate
FROM stagesnfportalpm.Proposal Proposal
JOIN stagesnfportalpm.Call ON Call.Id = Proposal.CallId
JOIN mdm.StateDatahub ON StateDatahub.StateDatahubId = '90283A0
WHERE
-- Cancelled Status --

```



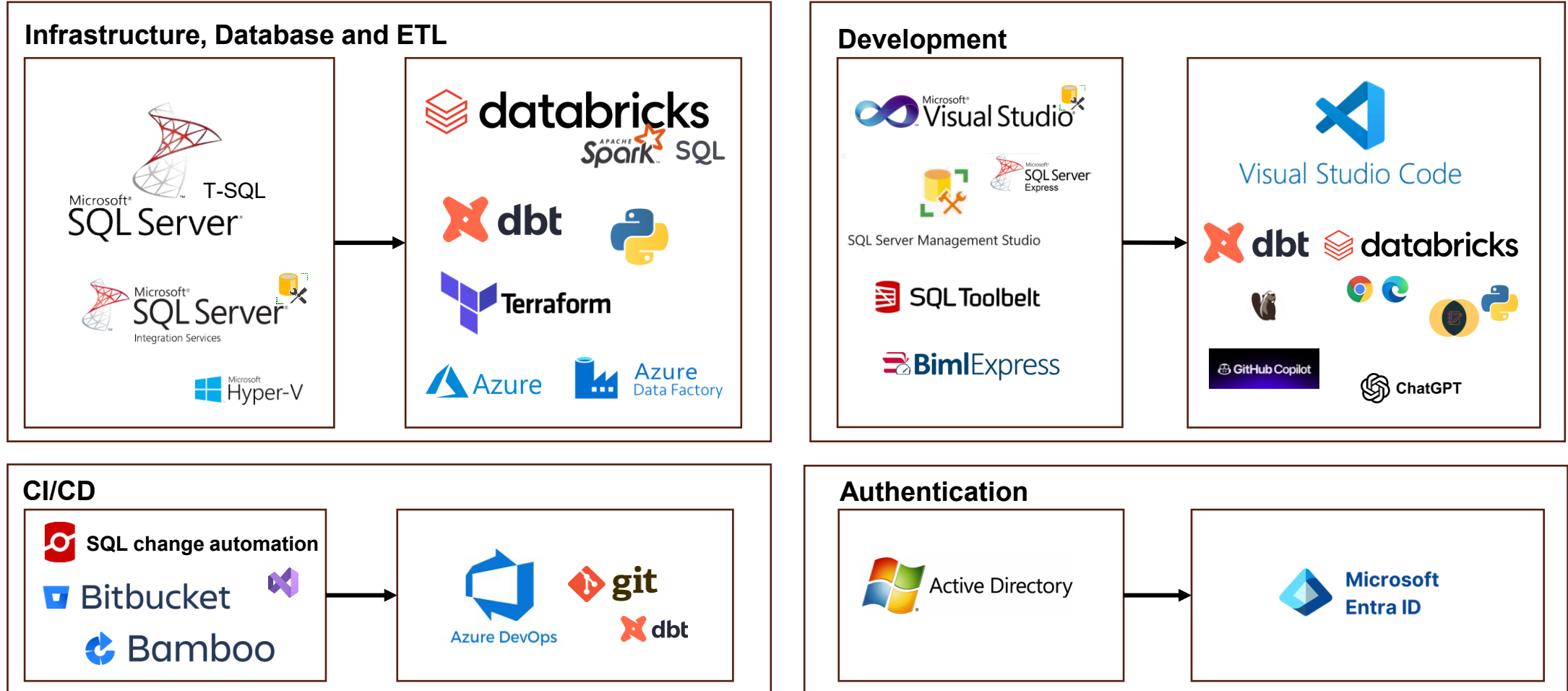
```
e_manifest_grouped.py IM app_Address.sql
datahub_lakehouse > src > models > Datahub > app > app_Address.sql
Add documentation or tests [Datapilot Notebooks | Olivier Humi, 4 months ago | 4 au
{{ config(alias='Address') }}
SELECT a.AddressId,
a.PersonBaseId,
a.ValidFrom,
a.TypeUid,
a.TypeSortNumber,
a.Type,
a.Type_De,
a.Type_Fr,
a.Street,
a.Supplement1,
a.Supplement2,
a.PostCode,
a.Place,
a.Canton,
a.CountryId,
c.IsoCode AS CountryIsoCode,
rCountry.English AS Country,
cast(a.IsCurrent as boolean) as IsCurrent,
cast(a.IsActive as boolean) as IsActive,
GETDATE() AS UpdatedInHDate
FROM {{ ref('Ga_Address') }} a
LEFT JOIN {{ ref('MySnf_Country') }} c ON c.CountryId =
LEFT JOIN {{ ref('MySnf_Resource') }} rCountry ON rCoun
UNION
/* person not imported address (only the ones with PersonId
SELECT pni.PersonNotImportedId AS AddressId,
pni.PersonNotImportedId AS PersonBaseId,
NULL AS ValidFrom,
addrType.TypeUid,
addrType.TypeSortNumber,
addrType.Type,
addrType.Type_De

```

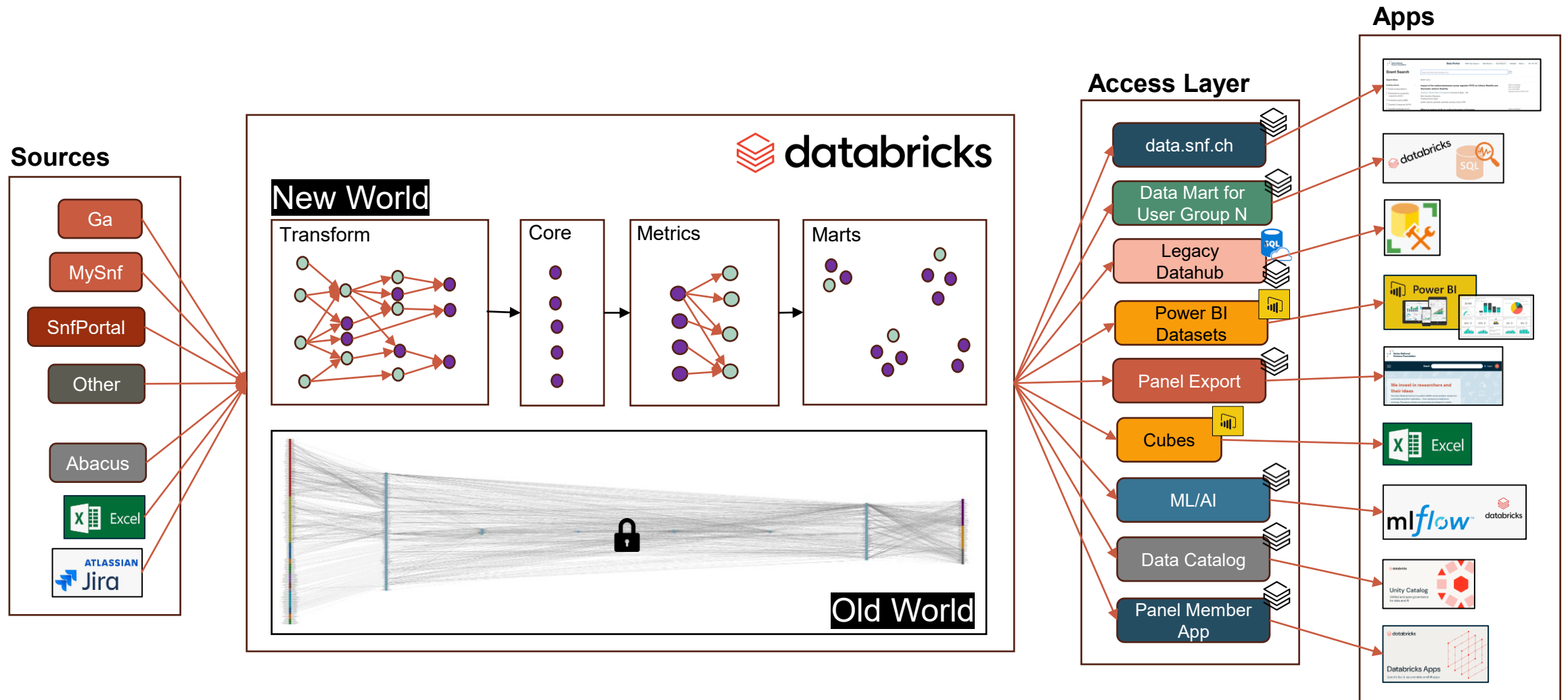


# Modernization overview

*“This wasn’t just a tech upgrade. It changed how our team works, delivers, and thinks and how we interact with our customers.”*



# Current step: remodeling the complete DWH

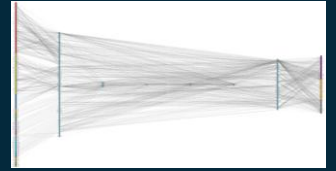


Open-source  
community

Databricks apps are  
super cool

Easier and more fun  
to onboard people

Better  
insights



Transparent user  
activities

Efficient operations

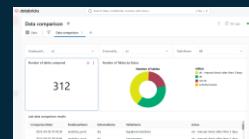
Tracking of  
dependencies

# Databricks and dbt are very powerful

Code First  
approach

- automatization
- benefit from AI

Custom tools for  
testing



Best practices from SW  
development

Don't be perfect from the  
start

Only start with  
Terraform after an  
MVP

Biggest challenges are  
not technical

Don't use legacy tech in  
the cloud

## Cloud set up is hard

Bottlenecks can delay  
the whole project

Collab with central PLAT  
team extremely important

Private end-points make  
everything more difficult

Change your strategy  
as you learn new facts

No one in our team  
had experience with  
the new Stack

Technology has no  
intrinsic value

## Be curious and learn fast

Embrace agile

Generative AI changed  
everything

Start small, fail fast

The new dev toolchain  
changed the way our  
team works

Thank you!