INSPARI

a valantic company

SAP to Databricks

Making the data journey work

Hosted by:

Daniel Størup Thomsen, Senior Data Engineer, Inspari

Peer Woyczechowski Bilberg, Senior Data Scientist, Inspari

Mikkel Hyldig, Data & Analytics Manager, Søstrene Grene

Today's Presenters

DANIEL STØRUP THOMSEN

Senior Data Engineer, Inspari

PEER WOYCZECHOWSKI BILBERG

Senior Data Scientist, Inspari

MIKKEL HYLDIG M. CHRISTENSEN

Data & Analytics Manager, Søstrene Grene

Practical Information

The webinar is being recorded. Recording and slides will be sent after the session.

Q&A will be held at the end of the presentation.

INSPARI a valantic company

SØSTRENE GRENES ALL OVER THE WORLD HANDELSKOMPAGNIE

Your Partner for Business Impact with AI & Data

Our mission

We empower **decisions, actions, and change** through people and data

Strong team

200+ data & AI experts

Copenhagen, Aarhus & Bucharest

Comprehensive capabilities for successful AI & data transformation

- Data & AI Strategy
- Data Architecture
- Data & AI Products
- Data Management & Governance
- Data Engineering
- Data & AI Ops

Data Value Guaranteed

Recognized thought leader

Podcast on data, AI, and digital transformation

Yearly conference with 500 attendees

Webinars & events

Microsoft MVPs

A well-built Microsoft proficiency

Fabric Featured Partner

Cloud Solution Provider

Azure

Microsoft Solutions Partner
Digital & App Innovation
Azure

Microsoft Solutions Partner
Business Applications

Microsoft Solutions Partner
Infrastructure
Azure

Microsoft Solutions Partner
Data & AI
Azure

Specialist
Analytics
AI and Machine Learning

Substantial track record of successful AI & data implementations

PANDORA · ENERGINET · Dansk Industri · BESTSELLER · SIEMENS · CPH · midt regionmidtjylland · bankdata · nne novo nordisk engineering

Selection of technology partners

snowflake · databricks

INSPARI
a valantic company

Søstrene Grene in Numbers

360
Stores globally

500+
HQ Sisters

52
HQ Sisters

15
Different countries

2.9b
DKK SG Group Turnover 24/25

12
Web shop markets

Agenda

- Challenge, Decision, Implementation @ Søstrene Grene
- Dataplatform Architecture
- Logic in SAP vs. Databricks
- Demo of the Solution

From Navision to SAP – A Need for a New Architecture

Challenge – Decision – Implementation

Starting Point (Early 2024)

Navision ERP + Azure SQL DWH via ADF pipelines.

SAP S4 HANA greenfield implementation.

The Challenge

Existing setup not suited for SAP as main data source.

Need for a future-proof setup that could scale with our business.

INSPARI
a valantic company

SØSTRENE GRENES HANDELSKOMPAGNIE ALL OVER THE WORLD

Hybrid Architecture as Foundation

Challenge – **Decision** – Implementation

Considerations

Databricks vs Fabric.

SAP Datasphere vs CDC-connection vs ADF.

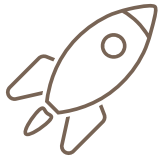
Decision

Databricks due to matureness and SAP partnership.

SAP Datasphere due to native integration and simplicity.

Development of Data Platform

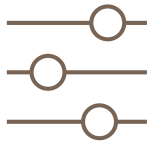
Challenge – Decision – **Implementation**



SAP Go-Live for Finance (October 2024)

Main components:

- Replication flows in SAP Dataphere
- DLT for data modelling in finance data models
- SAP Analytics Cloud build on top



SAP Go-Live 2.0 (Early 2026)

Expansion and adjustment of current setup to accommodate new reporting tracks.

Agenda

- Challenge, Decision, Implementation @ Søstrene Grene
- Dataplatform Architecture
- Logic in SAP vs. Databricks
- Demo of the Solution

Data Platform

The Architecture

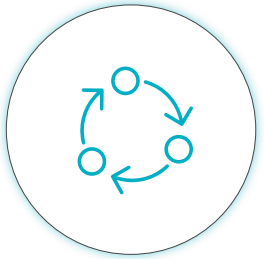
INSPARI
a valantic company

SØSTRENE GRENES HANDELSKOMPAGNIE
ALL OVER THE WORLD

Methodology behind the Data Platform

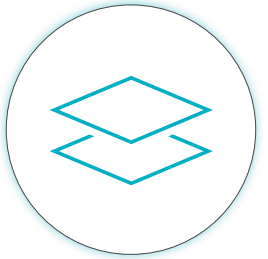
Scalability at the Core

The architecture is designed to seamlessly handle increasing data volumes and user demands. Horizontal scaling ensures consistent performance even as workloads expand across cloud-native components.



Automation-Driven Operations

End-to-end automation underpins pipeline orchestration, data quality enforcement, and CI/CD integration. This reduces human error, accelerates deployment cycles, and ensures governance at scale.



Flexibility Across Layers

Modular components and API-first design allow integration with heterogeneous tools and evolving business needs. Supports multi-cloud deployments, hybrid models, and vendor-agnostic infrastructure.

Technology Stack

The Foundation of the Data Platform

Databricks is the core platform that is used for its capabilities in Data & AI.

The platform makes it easy to work with structured, semi-structured, and unstructured data.

By utilizing structured streaming, it is possible to increase the data frequency without changing anything in the technical setup.

Datasphere replication flows facilitates data ingestion from SAP to Databricks.

Datasphere replication flows facilitates data ingestion from SAP to Databricks.

dbt handles the transformation layer in the modern data pipelines.

The Data engineers write SQL statements, and dbt converts them into tables and views in Databricks

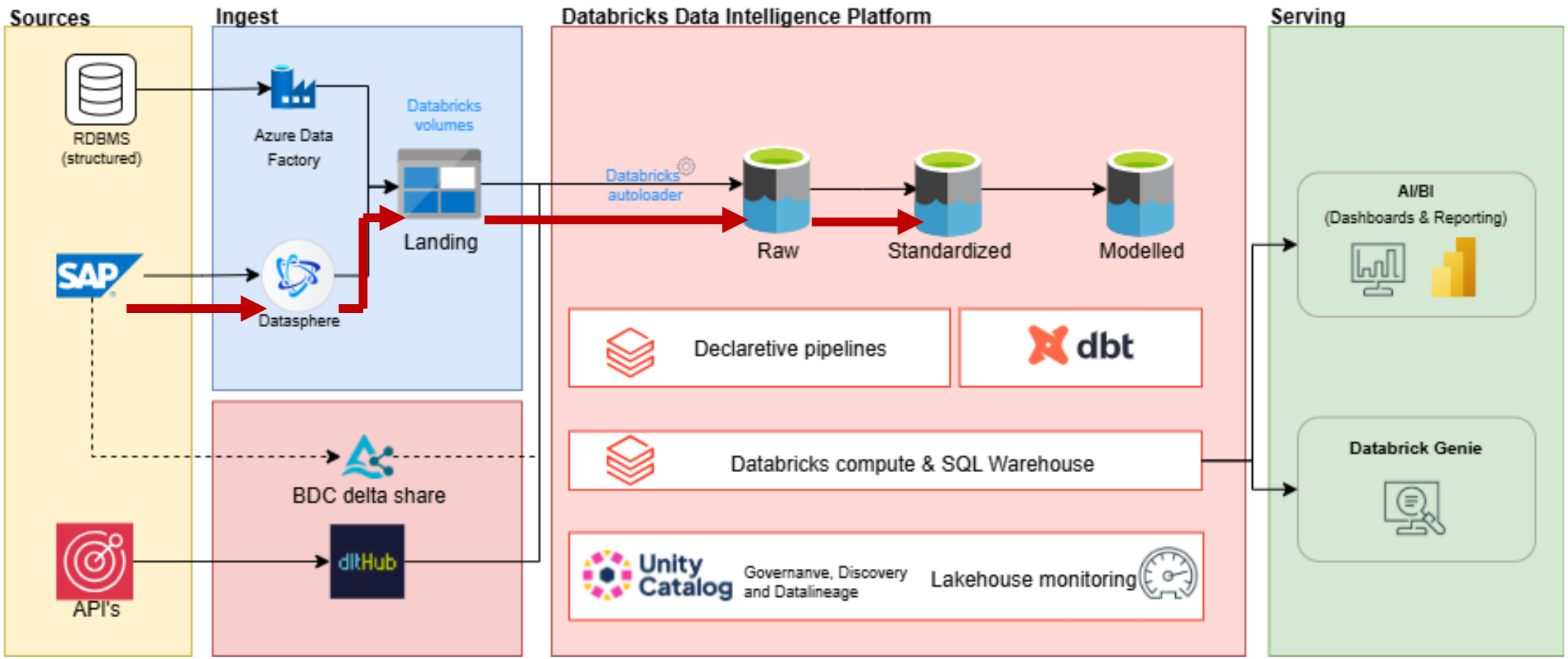
It brings software engineering practices like version control, testing, and documentation to data work. The modular approach lets teams build reusable data models that maintain consistency across transformations

Lakeflow Declarative Pipelines enables the creation of batch and streaming data pipelines using SQL and Python.

Users define data transformations, while the framework manages execution order and optimization.

Built on Spark, it features automated scaling, error recovery, and data quality monitoring. This approach allows data engineers to focus on business logic instead of pipeline infrastructure complexities.

The Architecture



Medallion Architecture

Bronze/ Raw Layer

Raw data ingested directly from SAP and other sources.

Stored in its original format for traceability.

Silver Layer

Cleansed and structured data.

Business rules applied, joins, and quality checks ensure reliable datasets.

Gold Layer

Curated, business-ready data marts and models.

Used for dashboards, analytics, and decision-making.

Data flows seamlessly from SAP and other systems into the **Bronze layer** → … is processed and validated in the **Silver Layer** → … and becomes trusted, business-ready information in the **Gold Layer**

Agenda

- Challenge, Decision, Implementation @ Søstrene Grene
- Dataplatform Architecture
- Logic in SAP vs. Databricks
- Demo of the Solution

Positioning Logic

Databricks vs. SAP Datasphere

databricks

Databricks (preferred for most transformations)

Centralize business logic in one platform → easier governance & reuse.

Scalable compute for heavy transformations (joins, aggregations, cleansing).

Flexibility to standardize across multiple sources, not just SAP.

Clear medallion layering (raw → standardized → modelled) supports traceability.

SAP

SAP Datasphere (CDS Views – when it makes sense)

Pushdown of very SAP-specific logic that is hard to replicate outside.

Reuse of existing business semantics already modelled by SAP teams.

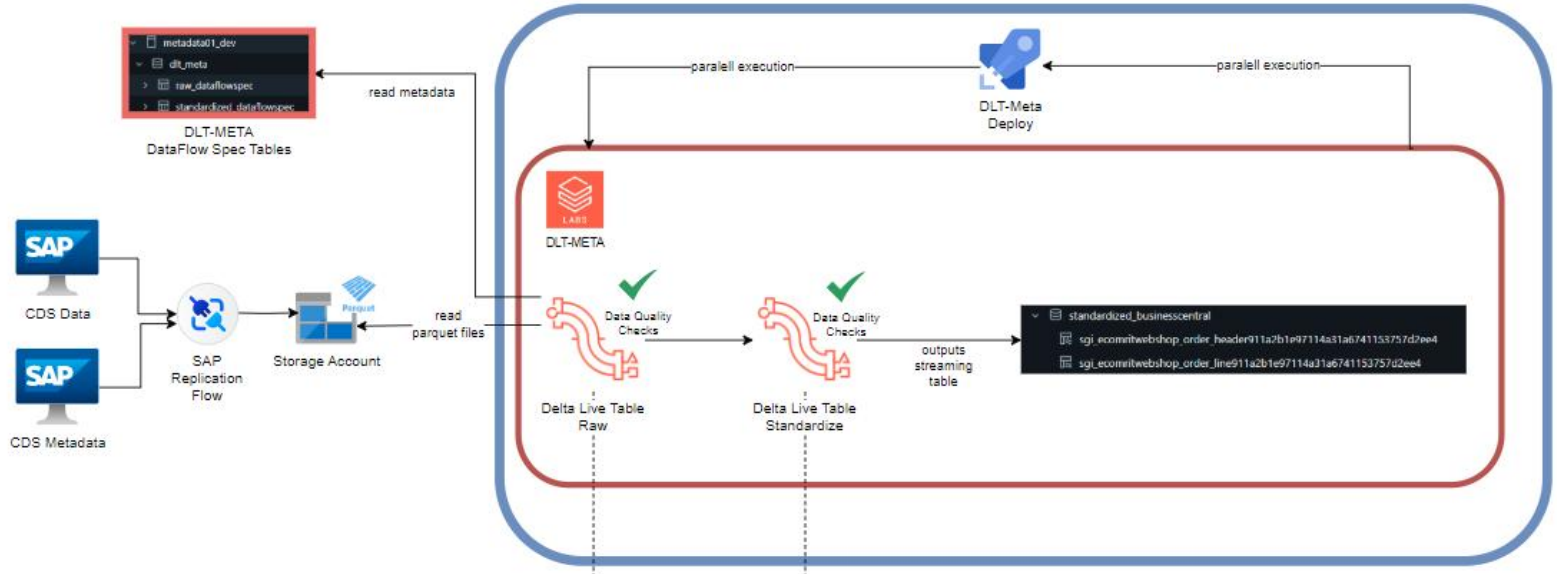
Reduce data volume early by filtering/aggregating before extraction.

Metadata Flow for Pipeline Generation

Meta-driven ingestion:
Seamlessly pull raw SAP data into a standardized Databricks layer.

Automated data delivery:
Free up engineers with an automated developer experience – focus energy on **business value**.

Streaming-ready by design:
Instantly enable real-time data flows with **Databricks Autoloader** if needed.



Agenda

- Challenge, Decision, Implementation @ Søstrene Grene
- Dataplatform Architecture
- Logic in SAP vs. Databricks
- Demo of the Solution

Demo

INSPARI
a valantic company

SØSTRENE GRENES
ALL OVER THE WORLD
HANDELSKOMPAGNIE

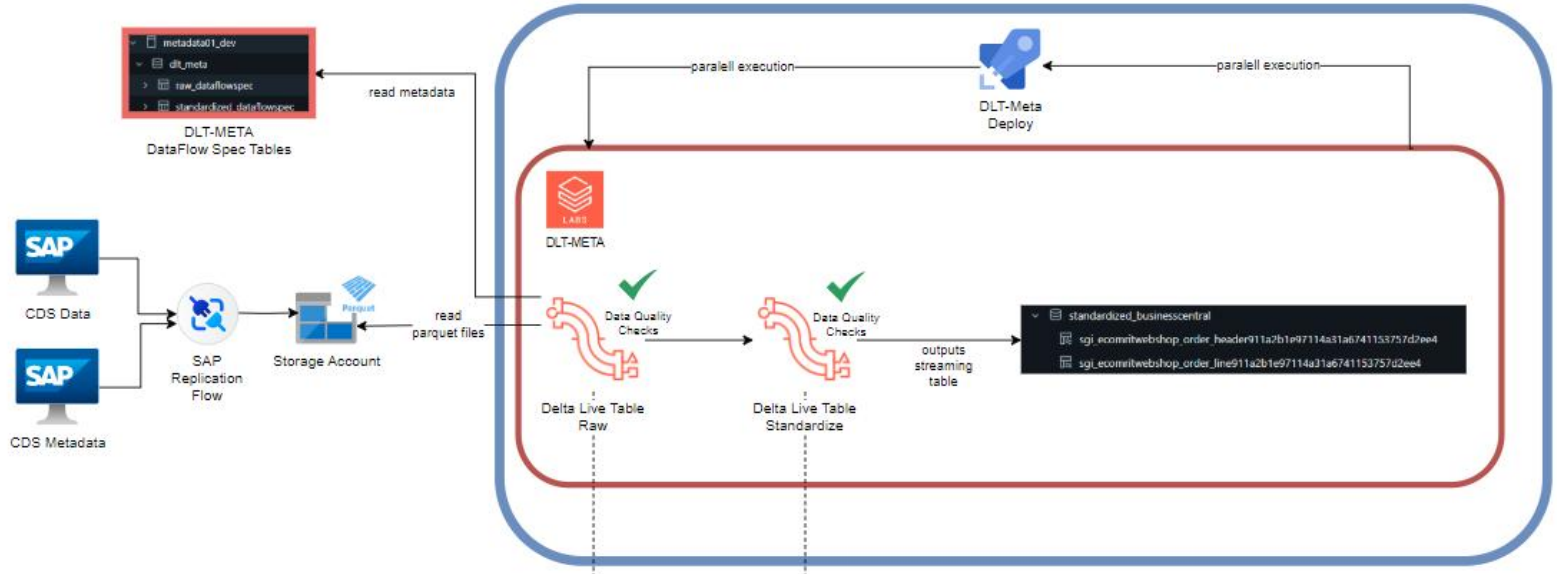
Metadata Flow for Pipeline Generation

Meta-driven ingestion:

Seamlessly pull raw SAP data into a standardized Databricks layer.

Automated data delivery:
Free up engineers with an automated developer experience – focus energy on **business value**.

Streaming-ready by design:
Instantly enable real-time data flows with **Databricks Autoloader** if needed.



Key Learnings

New tools, new roles

Do not underestimate the importance of clearly defined roles and responsibilities across BI, IT, end-users etc.

The value of MVPs

SAP can be a rabbit hole of endless possibilities (and tables). If possible, deliver the minimum viable product at first and built from there.

Not only a BI-project

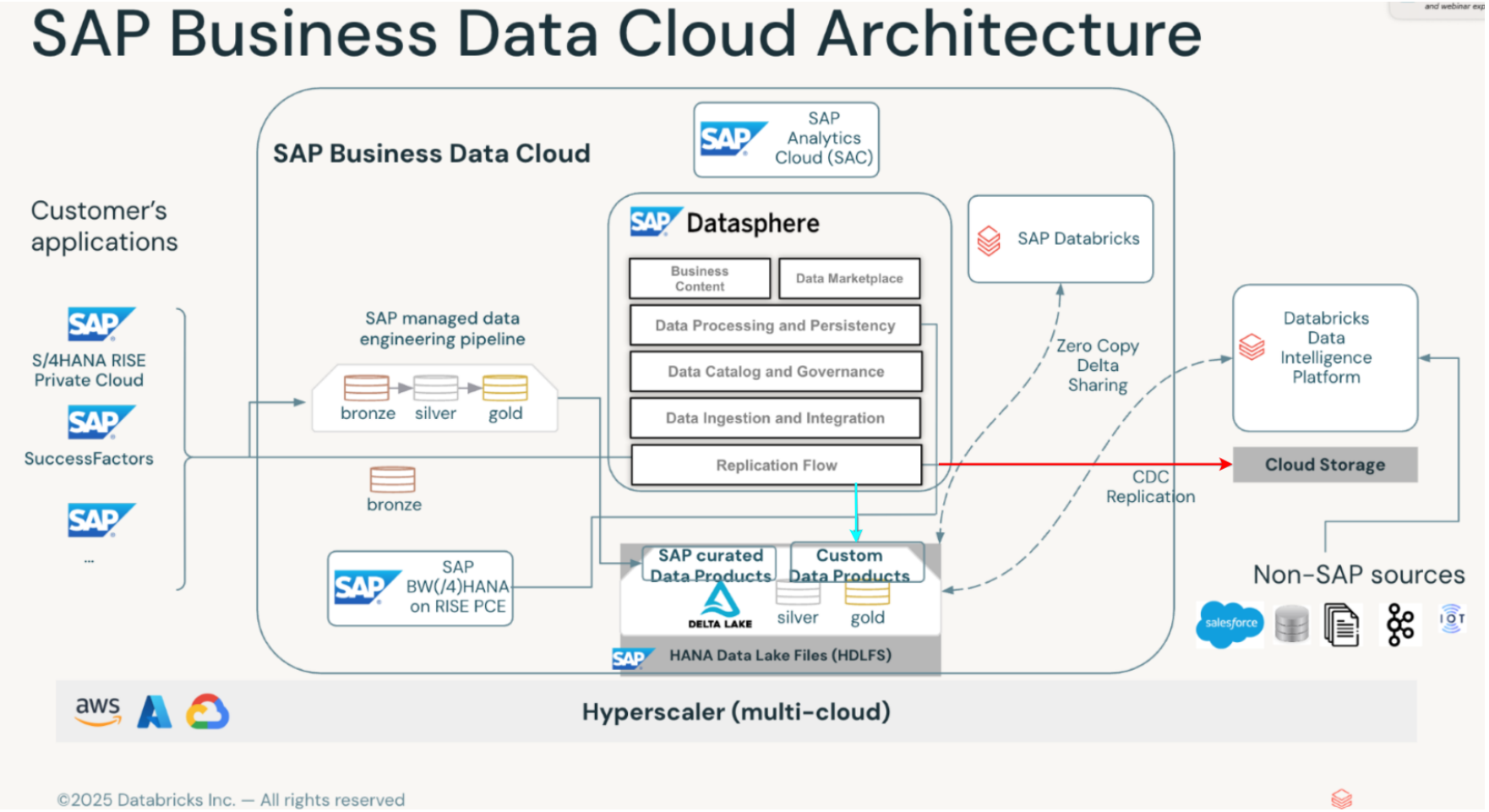
Success requires broad engagement; IT (SAP expertise), C-level (sponsorship) and end-users (clear, well-defined business needs).

SAP – Business Data Cloud x Databricks Connector

Coming to Azure Databricks in the beginning of October

Zero Copy Delta Sharing

SAP curated & custom data products



Time for Q&A

Get in touch!

In case you have questions, reflections, or seek sparring on something specific, please don't hesitate to reach out.

Daniel Størup Thomsen

dtho@inspari.dk

Peer Woyczechowski Bilberg

pewo@inspari.dk

Mikkel Hyldig M. Christensen

mikkel.hyldig@sostrenegrene.com

INSPARI a valantic company

SØSTRENE GRENES HANDELSKOMPAGNIE ALL OVER THE WORLD

Thank you.