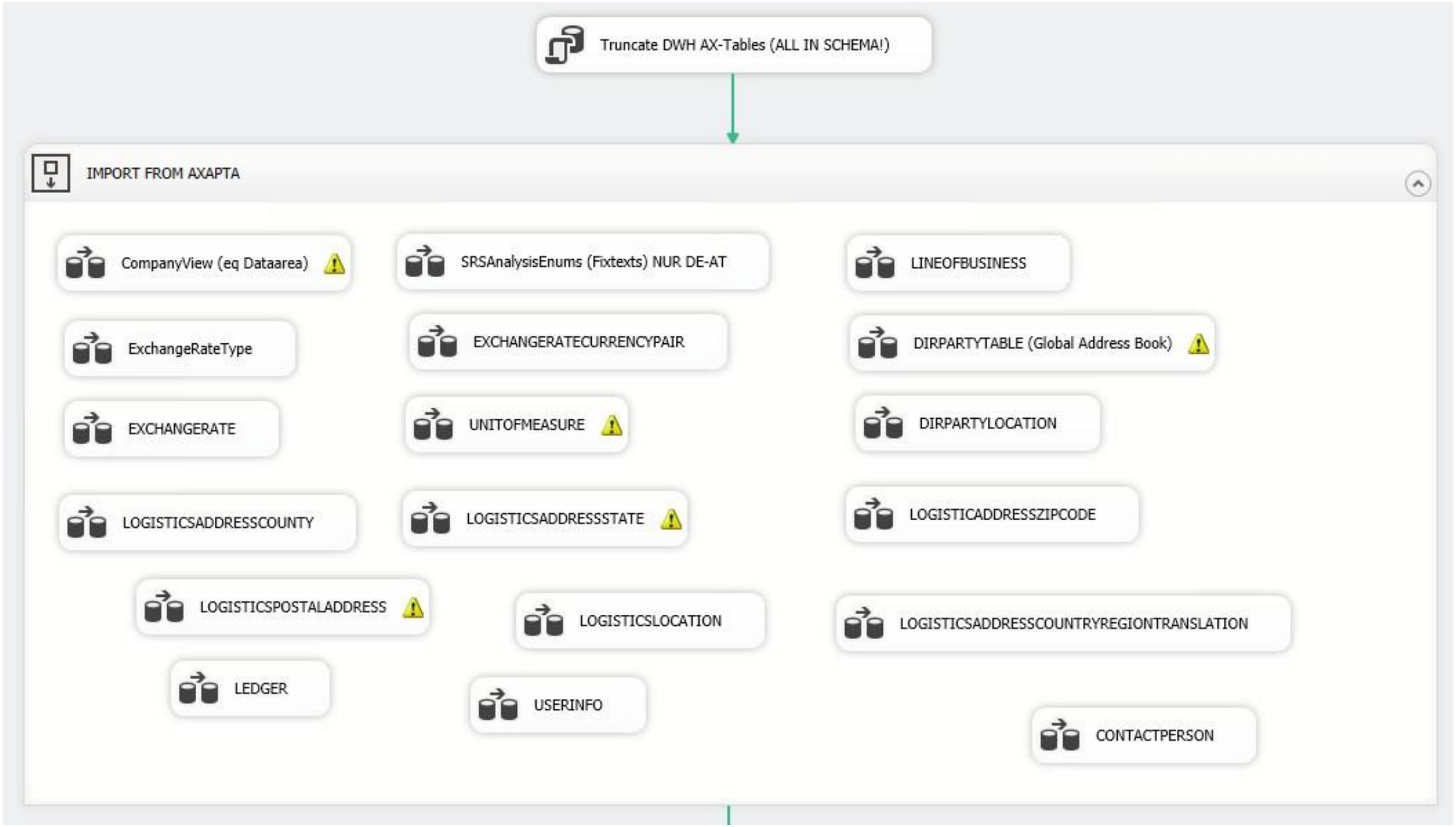
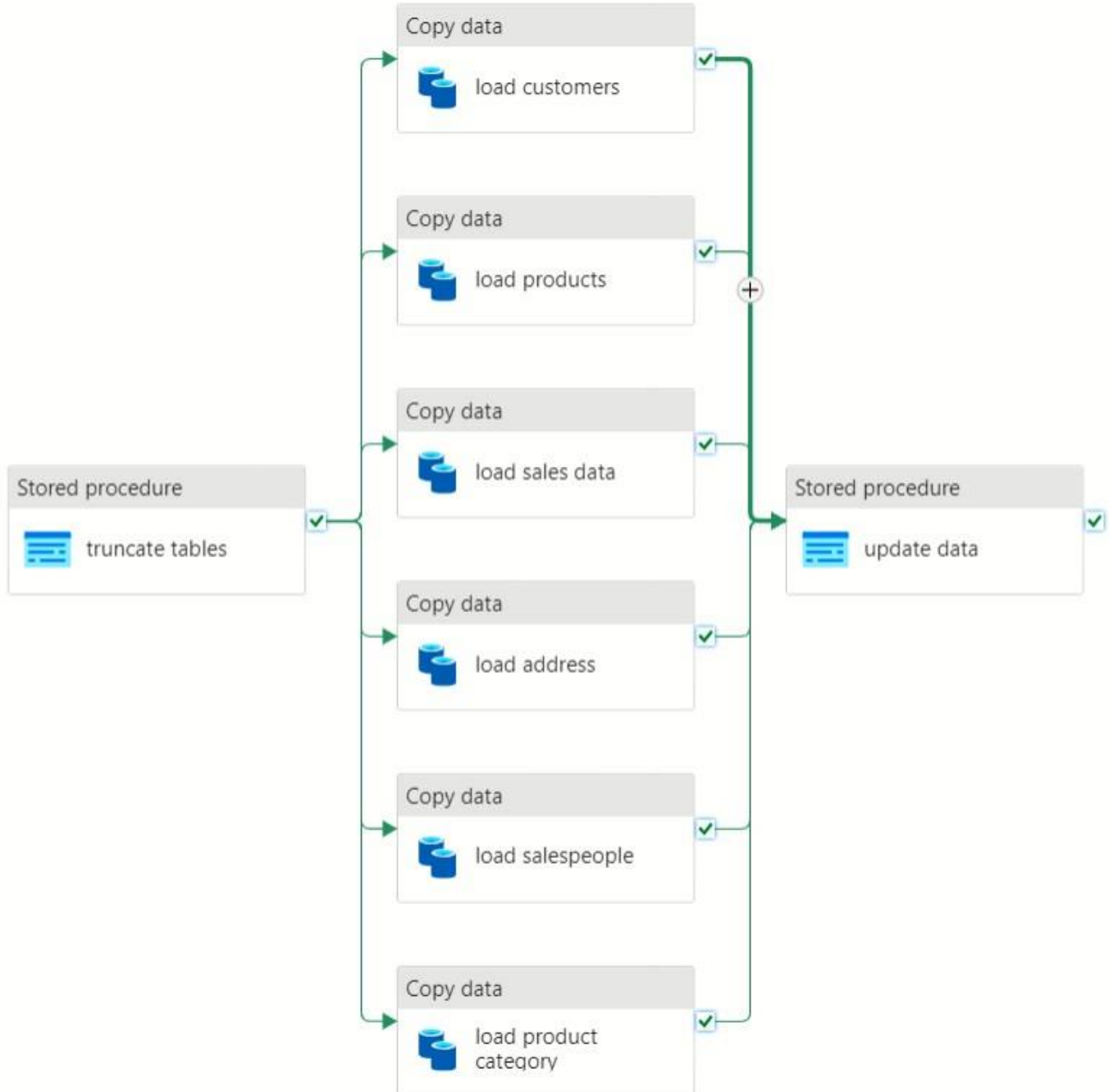
Speed up your development with metadata-based ETL processes

Bernhard Engleder
Teamlead Data & AI @ ACP CUBIDO digital solutions


Truncate DWH AX-Tables (ALL IN SCHEMA!)
IMPORT FROM AXAPTA
CompanyView (eq Dataarea)
SRSAnalysisEnums (Fixtexts) NUR DE-AT
LINEOFBUSINESS
ExchangeRateType
EXCHANGERATECURRENCYPAIR
DIRPARTYTABLE (Global Address Book)
EXCHANGERATE
UNITOFMEASURE
DIRPARTYLOCATION
LOGISTICSADDRESSCOUNTY
LOGISTICSADDRESSSTATE
LOGISTICADDRESSZIPCODE
LOGISTICSPOSTALADDRESS
LOGISTICSLOCATION
LOGISTICSADDRESSCOUNTRYREGIONTRANSLATION
LEDGER
USERINFO
CONTACTPERSON


Copy data
load customers
Copy data
load products
Copy data
load sales data
Stored procedure
truncate tables
Copy data
load address
Stored procedure
update data
Copy data
load salespeople
Copy data
load product
category

Back in the past: Hardcoded ETL

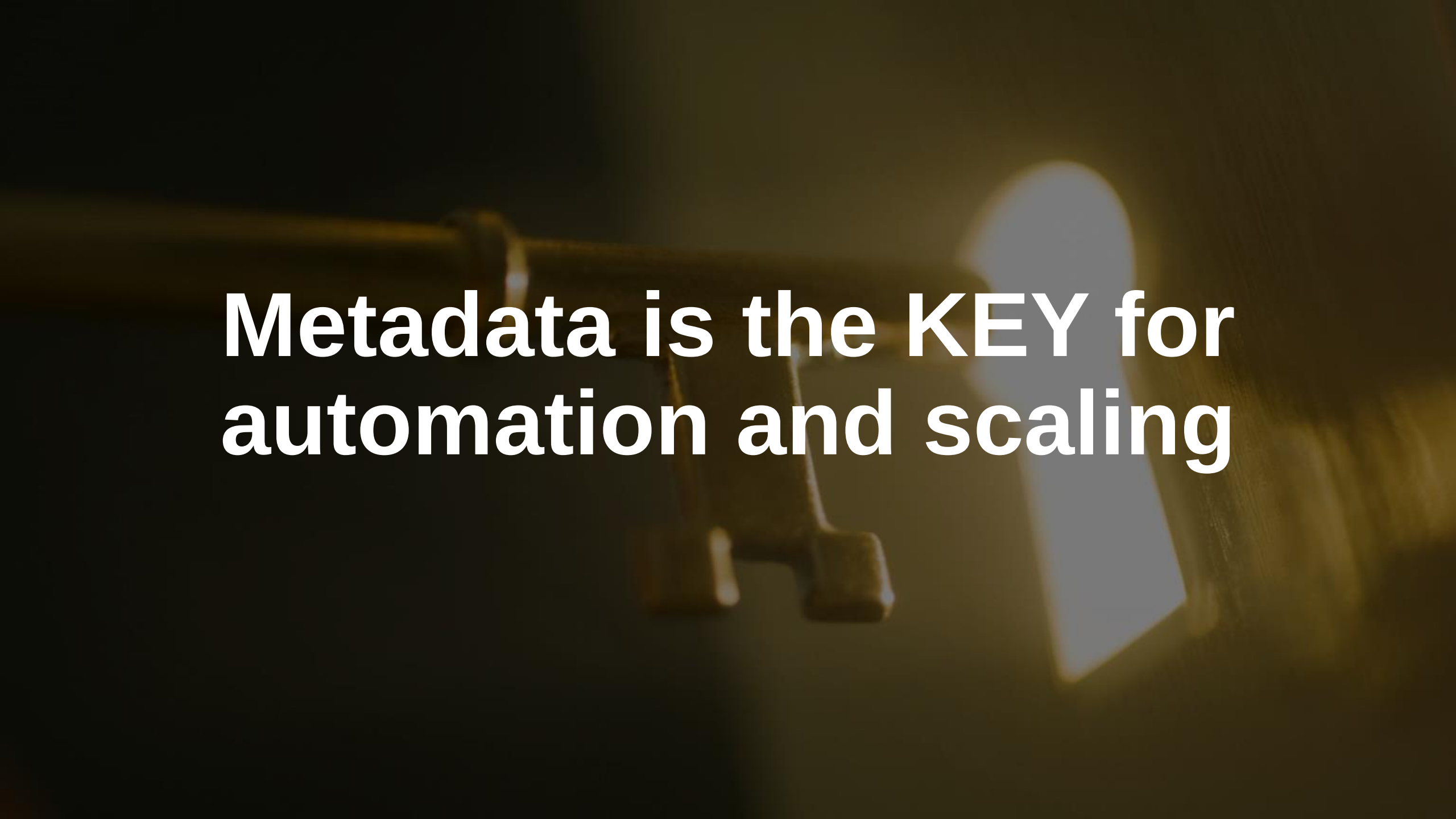
- For a table load (staging) we created each dataflow manually
- To load multiple tables from a source system to your staging area this results into many ETL objects (datasets/pipelines)



The old way: Problems / side effects of hardcoded ETL

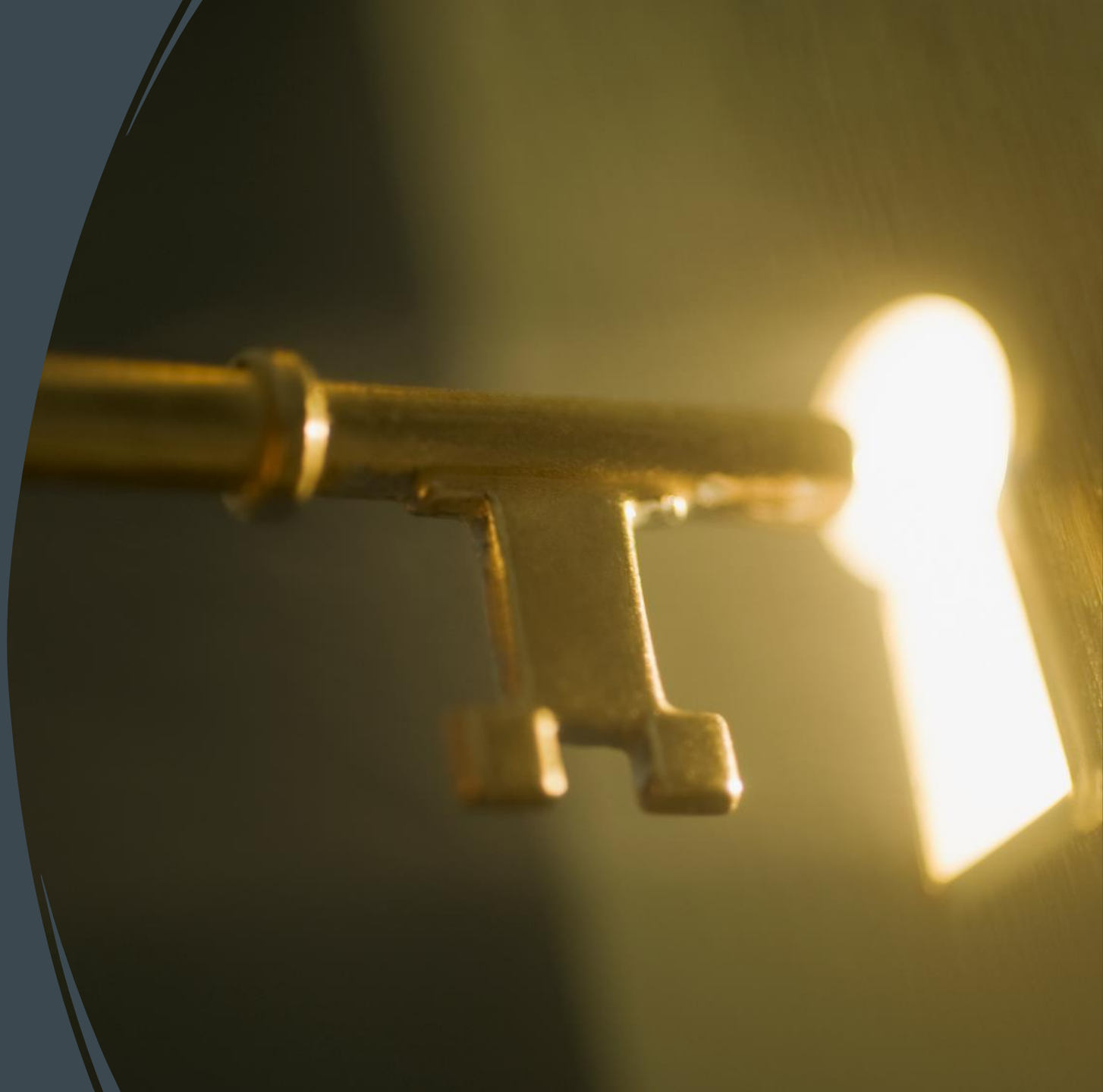
- High maintainance effort for bigger solutions
- Missing overview, lack of clarity
- Schema changes
- No scaling, same effort for each additional table
- Error rate during development




Metadata is the KEY for automation and scaling

Metadata is the KEY for automation and scaling

- Define and maintain the informations about your dataloads in central metadata tables
- Use this informations to automate your dataloads





(at least) 2 different approaches

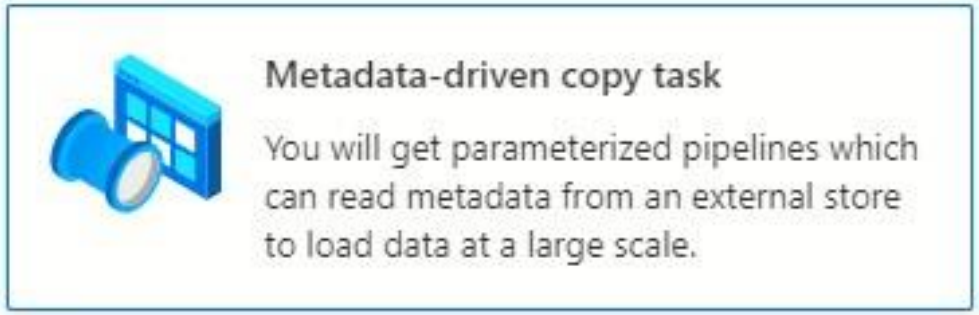
- Build you own configuration tables to
  - automate your dataload of multiple tables
  - Generate source code for your dataloads

- Use metadata based DWH development tools

IT for innovators.
Analytics & Development

Metadata-based pipelines in ADF/Synapse

- Templates available:
  - Metadata-driven copy task
  - Bulk Copy from Database

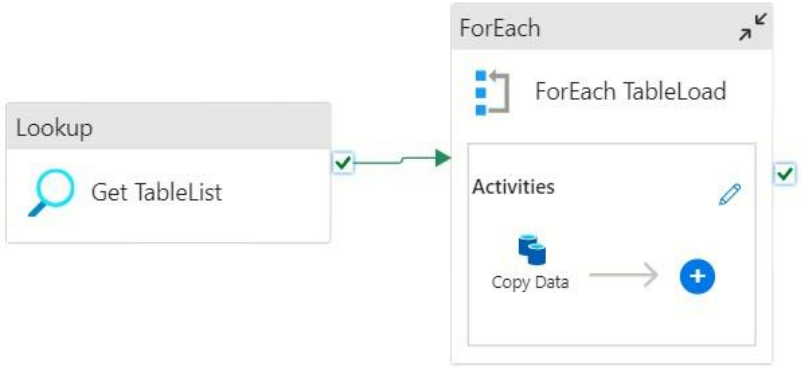




DEMO

Create your first metadata-based Pipeline

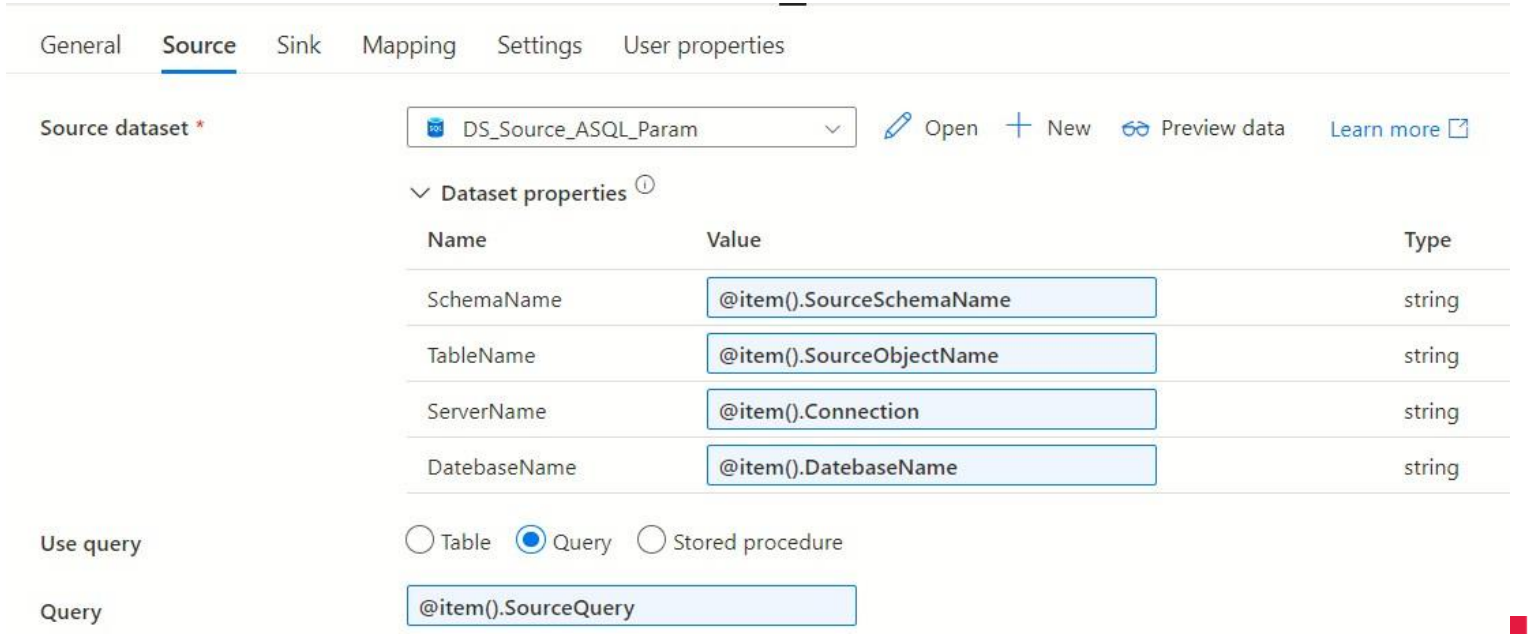
Metadata Driven ETL V1

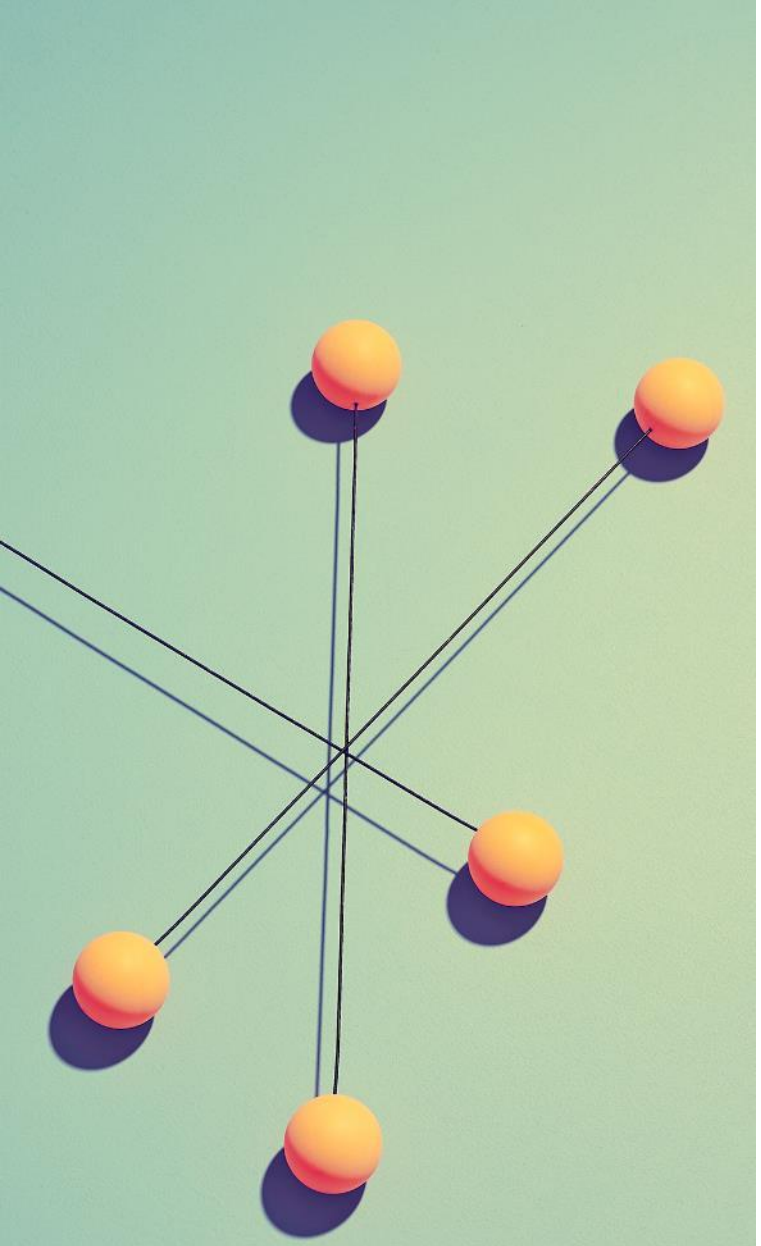




Parameters in ADF/Synapse pipelines

- Use parameters for
  - Copy data tasks
  - Pipelines
  - Procedures
  - Linked services





Enhance your configuration table

- Environments
- Scopes
- Selection List
- Business Keys
- Delta Load
- …

IT for innovators.
Analytics & Development

DEMO

Enhanced control table

Metadata Driven ETL V2





Enhance your metadata-based pipelines

- Pre-/Post-Copy-Tasks
- Logging
- Historization
- Delta Load



IT for innovators.
Analytics & Development



Generate source code based on metadata

- Generate procedures and index on demand or automatically
- Adapt your default procedures if necessary
- Use generated procedures to execute your dataload

IT for innovators.
Analytics & Development

DEMO

Enhance your metadata-based pipelines
Generate source code based on metadata

Metadata Driven ETL V3



Advantages

- Centralized administration of your dataload configuration
- Scaling: easy to add new tables
- React to schema changes automatically
- Reuse a pattern, lower risk of errors
- Naming conventions
- Documentation and traceability (Data Governance)



Metadata-based pipelines in ADF/Synapse

- Possibilities and levels of complexity are broad
- Caution: Don´t over-engineer your solution!
  - One solution to cover all your tasks and edge-cases can become difficult to maintain and understand
  - Your solution should save you time on development and maintenance
- Focus on the main cases and develop individual solutions for your edge-cases

IT for innovators.
Analytics & Development

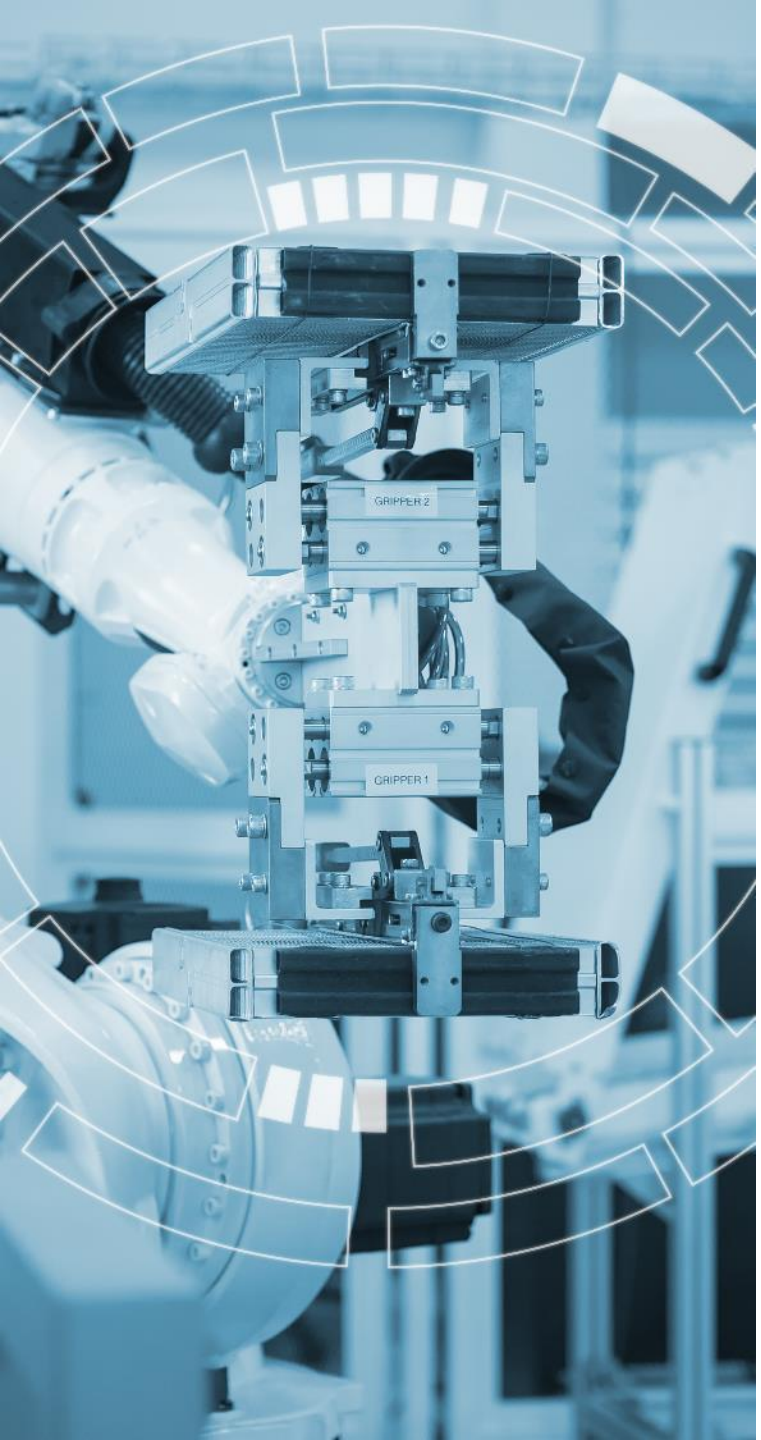
Metadata Driven Pipelines for Microsoft Fabric

Does the metadata-based pipeline from ADF/Synapse still work in Fabric?

➔ Yes, but…

- No template available yet
- Lookup not possible with query, only table
- No support for self-hosted integration runtime
- …





Tool support for metadata-based DWH

- Tools that use a central metadata datastore to develop, maintain and execute your metadata-driven ETL dataloads
- Reduce complexity and enable less-experienced data engineers to automate ETL

- Examples:
  - Cubido CuBIKit
  - TimeXtender

IT for innovators.
Analytics & Development

ACP CUBIDO Digital Solutions

Cubido CuBIKit

- Define Table and column mappings within a designer application and use a custom SSIS component to execute groups of dataloads

< | CuBIKit Designer | **Load** |

Filter

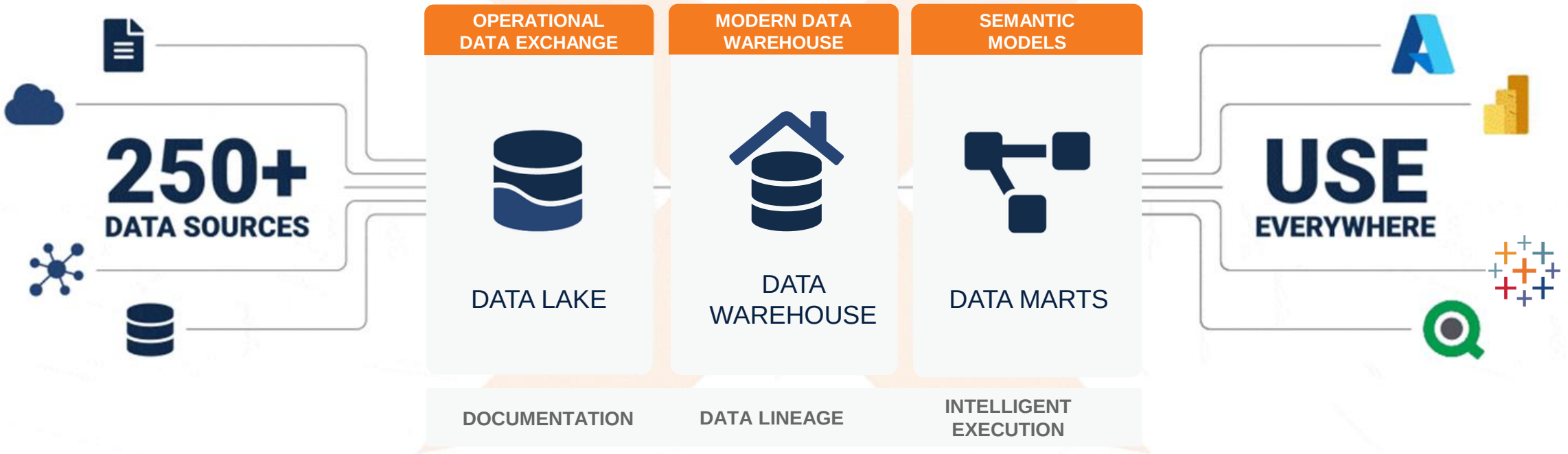
Table Mapping

Drag a column header here to group by that column.

Table Group	Source Schema	Source Table	Source Table Comm...	Source Table Locked	Destination Table Fil...	Destination Schema	Destination Table	Destination Table Co...	Target Table Locked
Einkauf	dwh	Bestellkopf	Bestellkopf	☐	DEFAULT	dbo	Bestellkopf	Bestellkopf	☐
Einkauf	dwh	Bestellposition	Bestellposition	☐	DEFAULT	dbo	Bestellposition	Bestellposition	☐
Einkauf	dwh	LieferdatumHistorie		☐	DEFAULT	dbo	LieferdatumHistorie		☐
Fakturenpositionen	dwh	Fakturenpositionen	View FPs	☐	DEFAULT	dbo	Fakturenpositionen	Fakturenpositionen	☐
Fakturenpositionen	dwh	Lieferbedingung	Lieferbedingung	☐	DEFAULT	dbo	Lieferbedingung	Lieferbedingung	☐
Fakturenpositionen	dwh	Tourenstandort	Tourenstandort	☐	DEFAULT	dbo	Tourenstandort	Tourenstandort	☐
Fakturenpositionen	dwh	Versandbedingung	Versandbedingung	☐	DEFAULT	dbo	Versandbedingung	Versandbedingung	☐
Fakturenpositionen	dwh	Lagerort		☐	DEFAULT	dbo	Lagerort		☐
Fakturenpositionen	dwh	GrundGutschrift		☐	DEFAULT	dbo	GrundGutschrift		☐
Fakturenpositionen	dwh	LagerbuchungAusgangsrechnung		☐	DEFAULT	dbo	LagerbuchungAusga...		☐
Kundenstamm	dwh	Kundenstamm	View Kundenstamm	☐	DEFAULT	dbo	Kundenstamm	Kundenstamm	☐
Kundenstamm	dwh	Debitorensaldo		☐	DEFAULT	dbo	Debitorensaldo	Debitorensaldo	☐
Kundenstamm	dwh	KundenOffenePosten		☐	DEFAULT	dbo	KundenOffenePosten		☐
Lieferantenstamm	dwh	Lieferantenstamm	View Lieferantensta...	☐	DEFAULT	dbo	Lieferantenstamm	Lieferantenstamm_T...	☐

IT for innovators. Analytics & Development

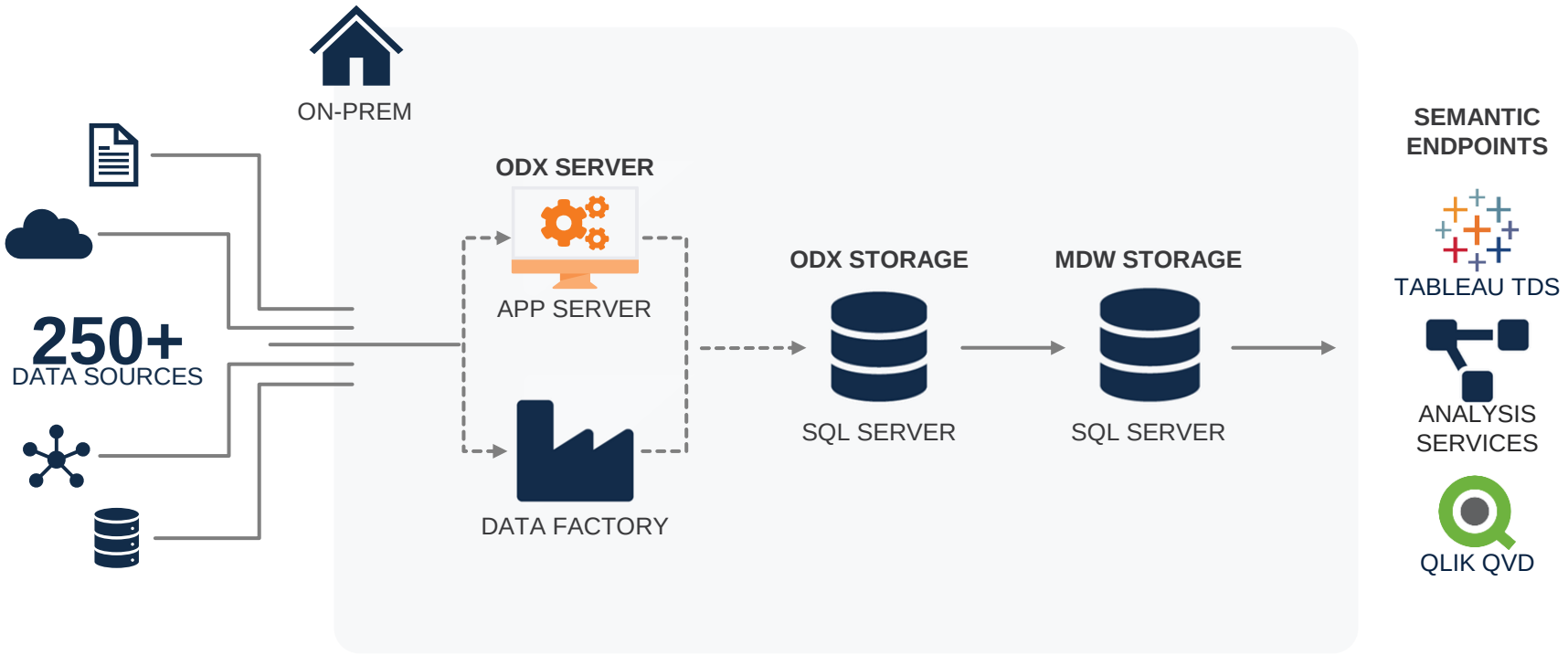
TIMEXTENDER



A Holistic, Metadata-driven Solution for **Data Integration**

On-Prem SQL Server

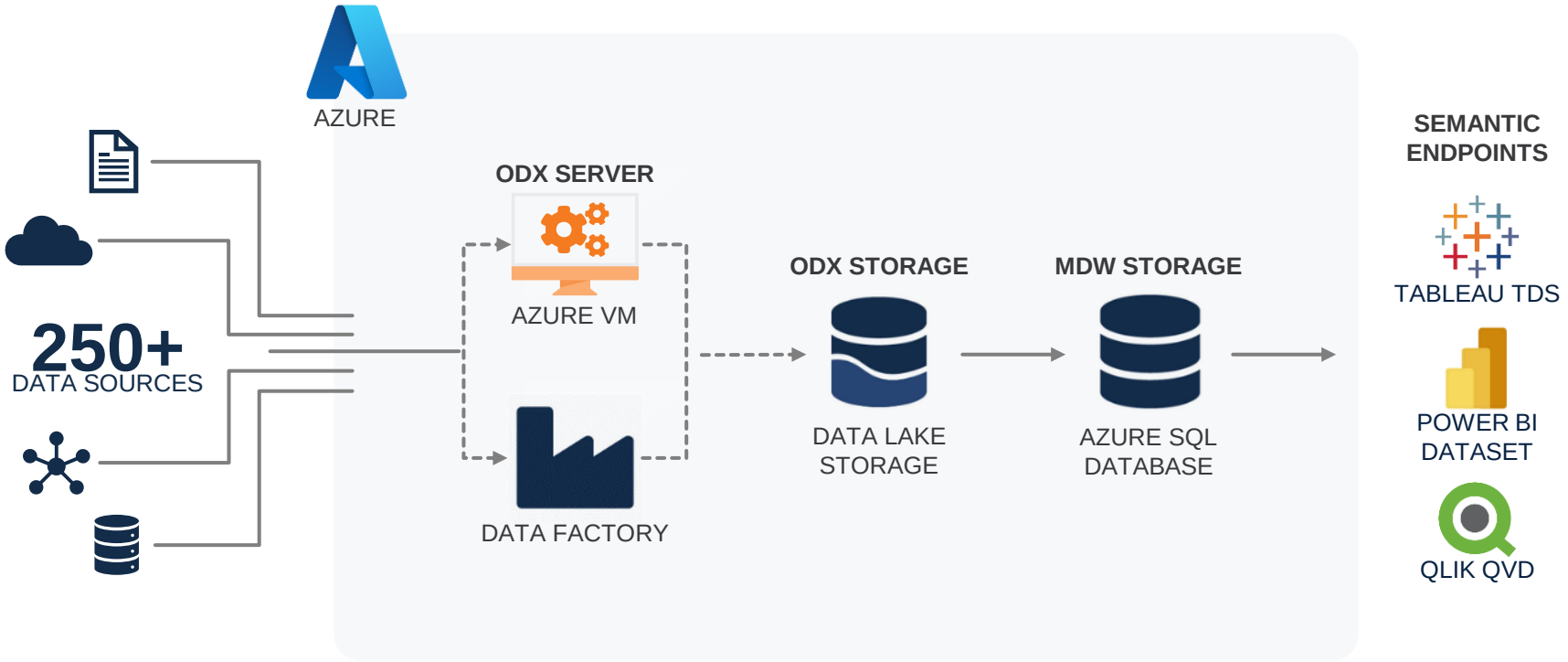
Reference Architecture



Futureproof your existing solution… Adapt to new technology without rebuilding

Azure SQL Database

Reference Architecture



Affordable solution for small-medium business reporting

DEMO

TIMEXTENDER





Key Takeaways

- Use metadata-based ETL to
  - scale your dataloads
  - introduce patterns
  - reduce risk of errors
  - Increase transparency and maintainability
  - lower the threshold for new data engineers in the project
- You can use individual developed pipelines in ADF/Synapse/Fabric when you have enough skilled data engineers and/or complex use-cases to solve
- Consider using a tool-based approach for your standard ETL processes to reduce complexity and the need for coding

IT for innovators.
Analytics & Development

Q&A

Thank you for your FEEDBACK!



sqlkonferenzeval.azurewebsites.net/211811570625952

Thank you for your attention!



Bernhard Engleder
Teamlead Data & AI
b.Engleder@cubido.at



Frank Wagner
Solution Specialist
fwagner@timextender.com

THANK YOU TO OUR SPONSORS

PASS Microsoft Data Platform Community DEUTSCHLAND e.V.

pmOne

SOLARWINDS

smart casual datadesign

Mainzer Datenfabrik

b.telligent
smart data. smart decisions.

HMS analytical software

ORAYLIS
turn your data into value

TIMEXTENDER

Microsoft

ITGAIN

SOLISYON
www.solisyon.de

GDS BUSINESS INTELLIGENCE GmbH



Thank you very much for your attention.

Vielen Dank für Eure Aufmerksamkeit.

ACP CUBIDO Digital Solutions

Aus Informationen Zukunft gestalten.