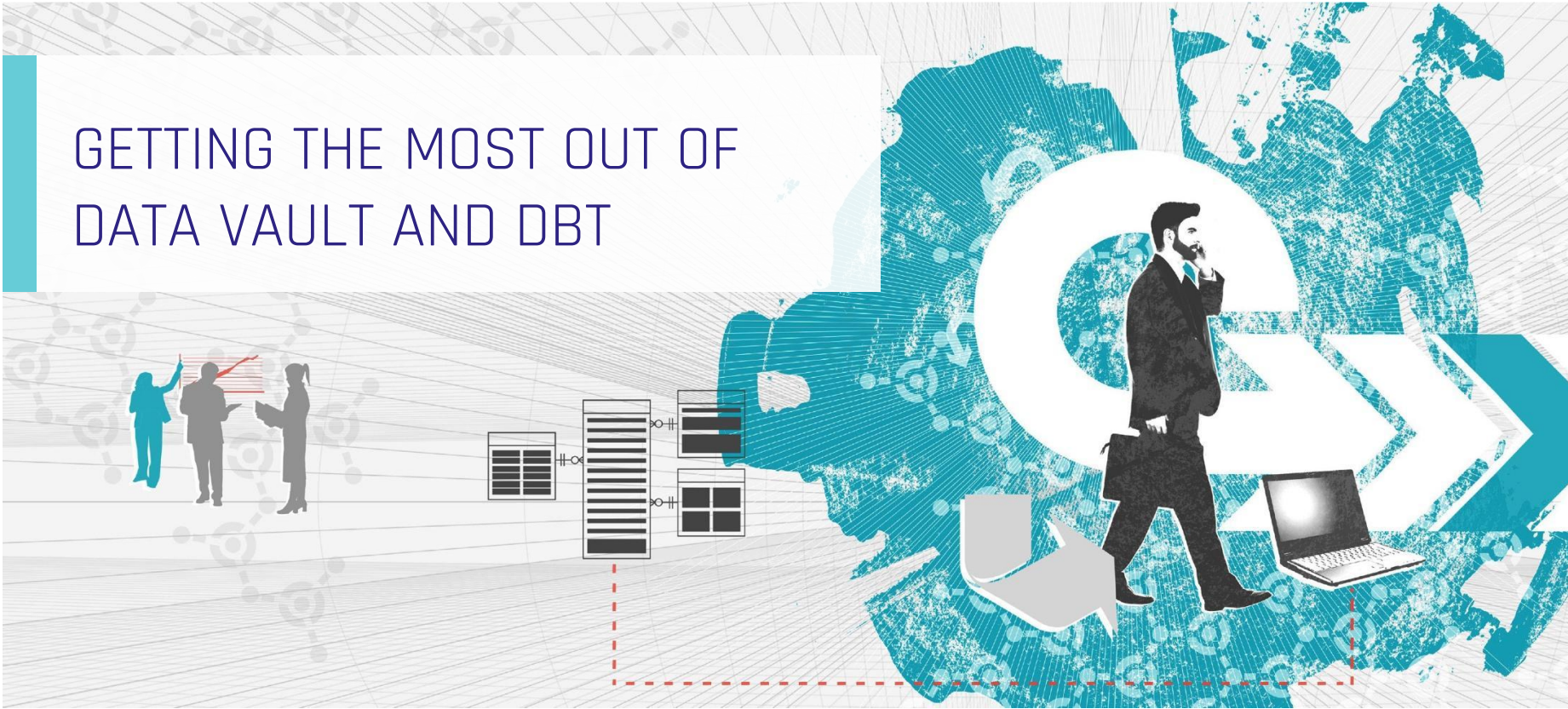


GETTING THE MOST OUT OF DATA VAULT AND DBT



AGENDA

1.

About us

2.

dbt in the Enterprise Data Warehouse Architecture

3.

Your typical Project Schedule with dbt

4.

Typical Challenges in Data Vault projects and how to avoid them

5.

Your Development Workflow with Turbovault & Datavault4dbt

6.

Demo Session with Datavault4dbt

OLE BAUSE

Technical Solutions Specialist (Scalefree International GmbH)



Experience

Two years of experience of Data Warehousing with Data Vault 2.0 implementation, Data Engineering and Data Warehouse automation.

Education

Master Applied Computer Science (major field Data Science)

Bachelor's degree in Business Information Systems (major field Business Intelligence)

Core Competencies

- Data Vault 2.0 modeling
- Data Transformation and Data Warehouse automation with dbt
- Data Engineering & ETL Development
- Cloud Computing



JONAS SUBEL

BI Consultant (Scalefree International GmbH)



Experience

Working with Scalefree on Data Vault 2.0 implementation, Data Engineering and Warehouse automation since August 2023

Education

Graduate of 42 Wolfsburg

Master of Science (Mechanical Engineering)

Bachelor of Science (Process Engineering)

Core Competencies

- Data Vault 2.0 modeling
- Data Transformation and Data Warehouse automation with dbt
- Data Engineering & ETL Development
- Cloud Computing



MICHAEL OLSCHIMKE



DAN LINSTEDT

TRAININGS

DATA VAULT CLASSES
AGILE WORKSHOPS
TOOL CLASSES



CONSULTING

ADVISORY
IMPLEMENTATION



EVENTS

DATA DREAMLAND
COMMUNITY EVENTS



KNOWLEDGE

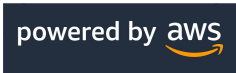
KNOWLEDGE ARTICLE
DATA VAULT FRIDAY
WEBINARS & SESSIONS
SOLUTIONS



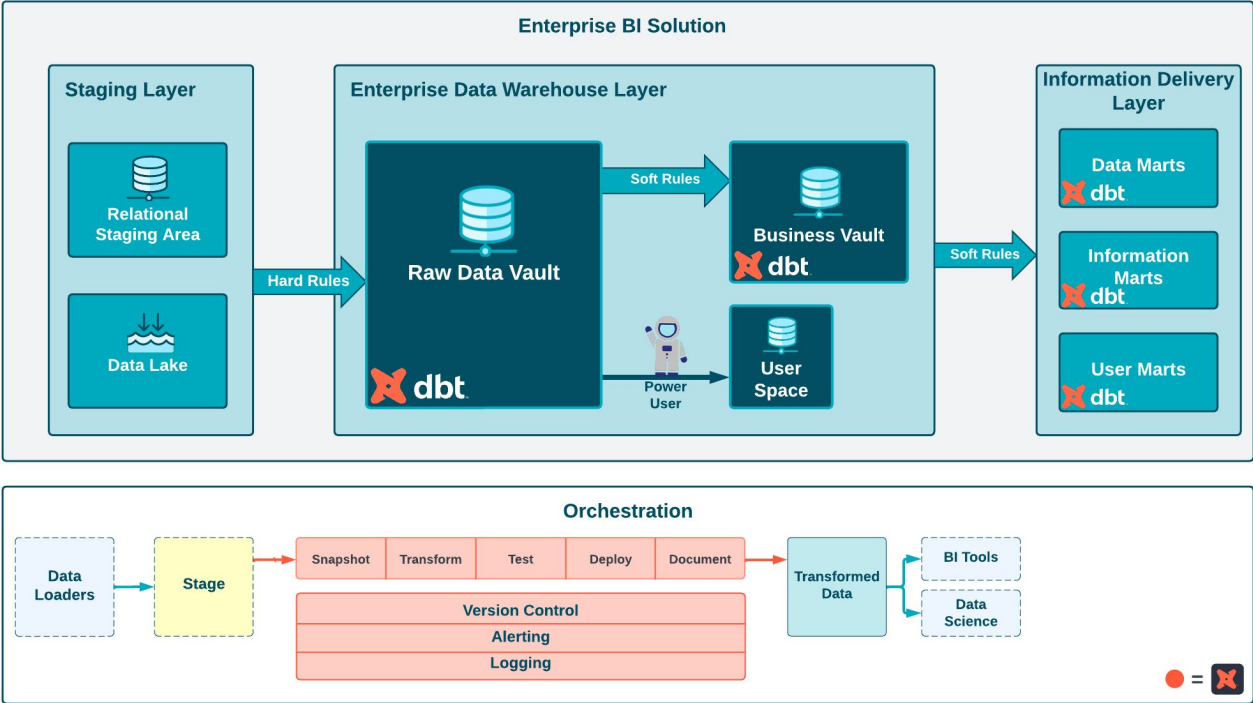
PARTNER SELECTION



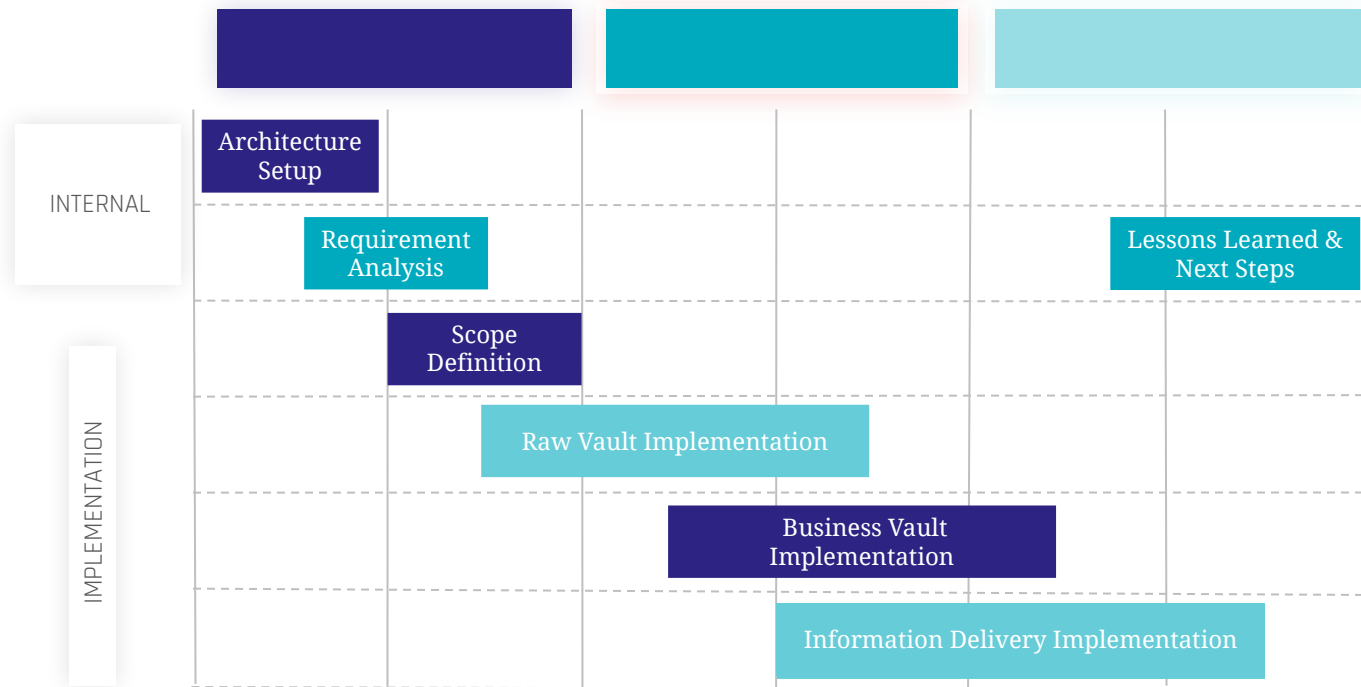
FURTHER SELECTED PARTNERS



DBT IN THE ENTERPRISE DATA WAREHOUSE ARCHITECTURE



YOUR TYPICAL PROJECT SCHEDULE



TYPICAL CHALLENGES

AUTOMATION LEVEL

There is still lots of manual work to do

DV STANDARDS?

Data Vault 2.0 standards are difficult to keep with too much manual work



POOR DATA VAULT MACROS

Own macros can be inefficient

- Customization required
- Performance issues

LOADING PATTERNS?

Standardized Loading Patterns are missing. Write every hub model yourself...

The background is a solid teal color. On the left side, there is a vertical light blue bar. In the center and right, there are several overlapping geometric shapes: a circle, a horizontal bar, and a large circular structure composed of concentric rings and segments, resembling a stylized gear or a complex diagram element.

HOW TO SOLVE THEM

OPEN SOURCE DBT PACKAGES FROM SCALEFREE





- dbt Package with Data Vault 2.0 patterns
- Fully auditable Data Warehouse solution congruent with Data Vault 2.0 definition
- Insert-only approach that avoids updating data
- Ready for both Persistent and Transient Staging Areas with multi-batch processing
- 25+ Global Variables
- Optimized incremental loads by implementing a high-water-mark that also works for entities that are loaded from multiple sources
- Based on years of experience



<https://hub.getdbt.com/Scalefree/eCOM/datavault4dbt/latest/>

CURRENTLY SUPPORTED PLATFORMS



UPCOMING



ORACLE



SUPPORTED DATA VAULT 2.0 ENTITIES

HUB

- Standard
- Reference

SATELLITE

- Standard (version 0 and 1 with virtualized LEDTS)
- Reference (version 0 and 1)
- Multi-Active (version 0 and 1)
- Record-Tracking
- Non-Historized

LINK

- Standard
- Non-historized

BUSINESS VAULT

- Point-in-Time Tables (with cleanup hook)
- Snapshot Control Tables
- Reference Tables

```
1  {{ config(materialized='incremental') }}
2
3  {% set yaml_metadata -%}
4  primaryKey: 'hk_account_h'
5  business_keys:
6    - account_key
7    - account_number
8  source_models: stage_account
9  {% endset -%}
10
11 {% set metadata_dict = fromyaml(yaml_metadata) -%}
12
13 {{ datavault4dbt.hub(hashkey=metadata_dict['hashkey'],
14                    business_keys=metadata_dict['business_keys'],
15                    source_models=metadata_dict['source_models']) }}
```

METADATA + MACRO = EXECUTABLE SQL

UNIQUE FEATURES OF DATAVAULT4dbt

POWERED BY SCALEFREE



**MULTI
BATCH**

Processing multiple batch loads from staging (PSA) into Raw Data Vault in one run by keeping a very good performance.

FLEXIBILITY

Many global variables configurable to adjust the package without changing the macros.

**SCALEFREE
SUPPORT**

Direct access to the development team of datavault4dbt.

**TURBOVAULT
4dbt**

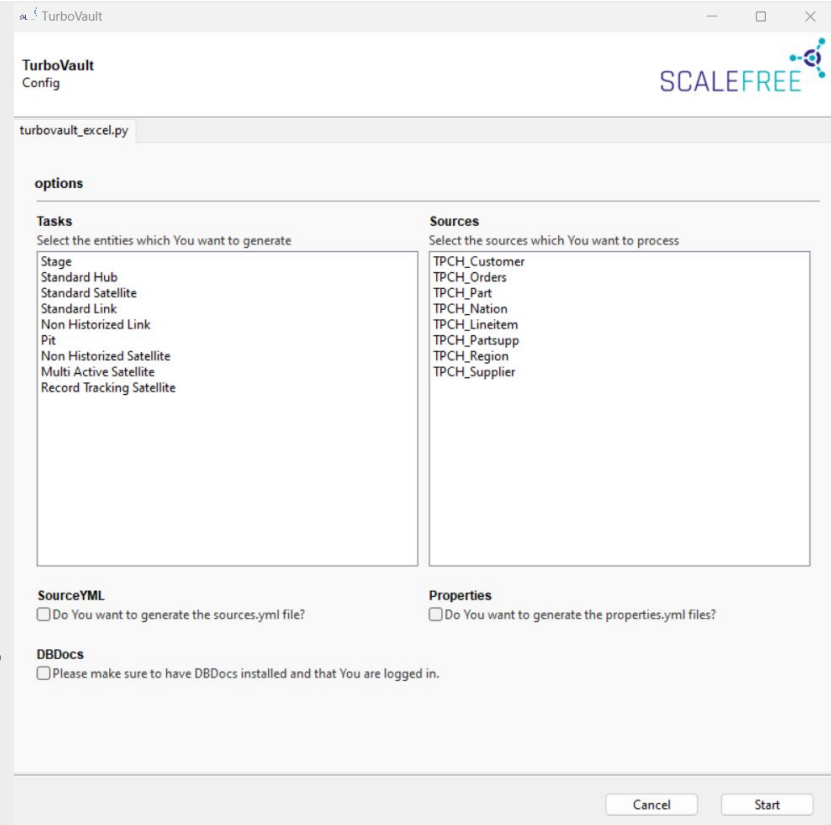
Generate dbt models for the Raw Data Vault based on structured metadata (no hand-written dbt models for the Raw Data Vault anymore!).



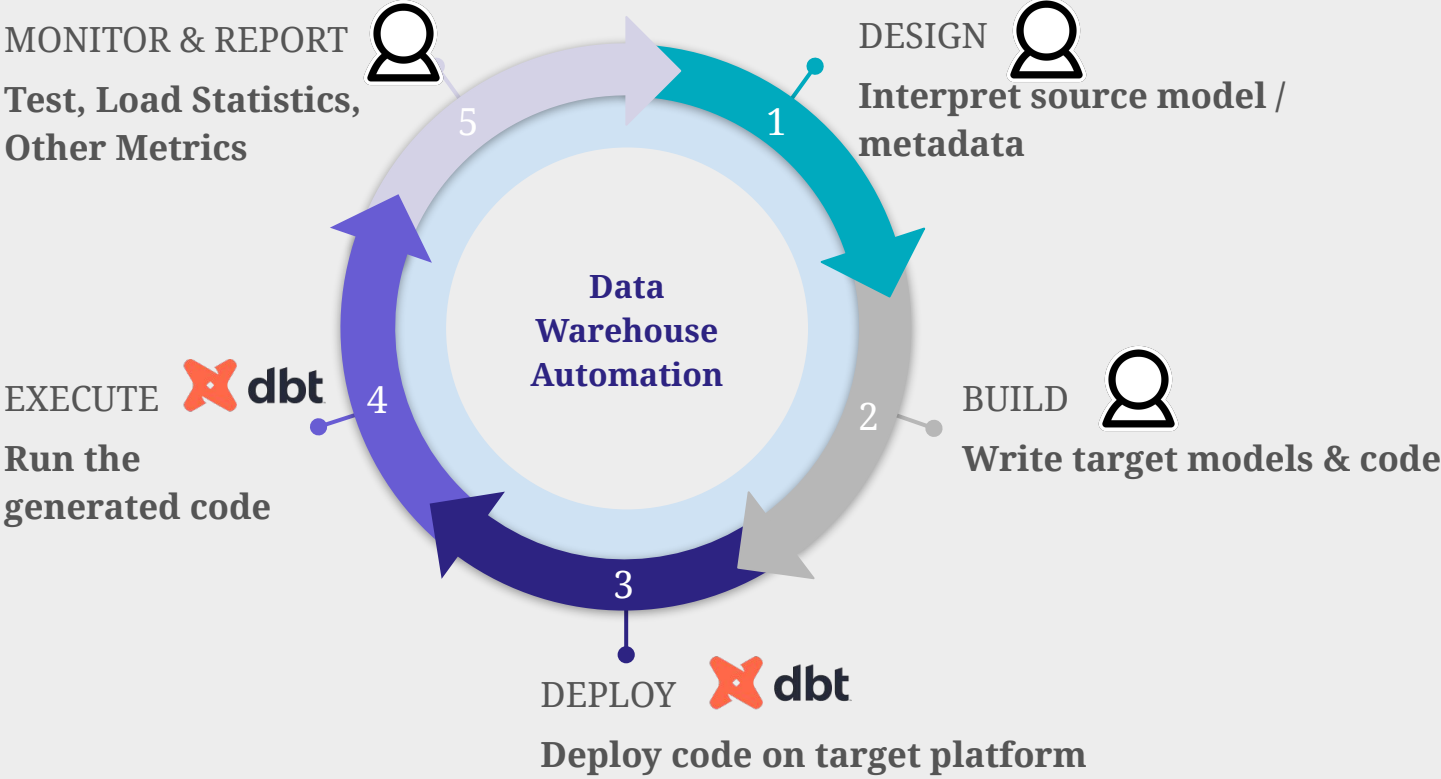
TURBOVAULT4dbt

Your dbt Model Generator

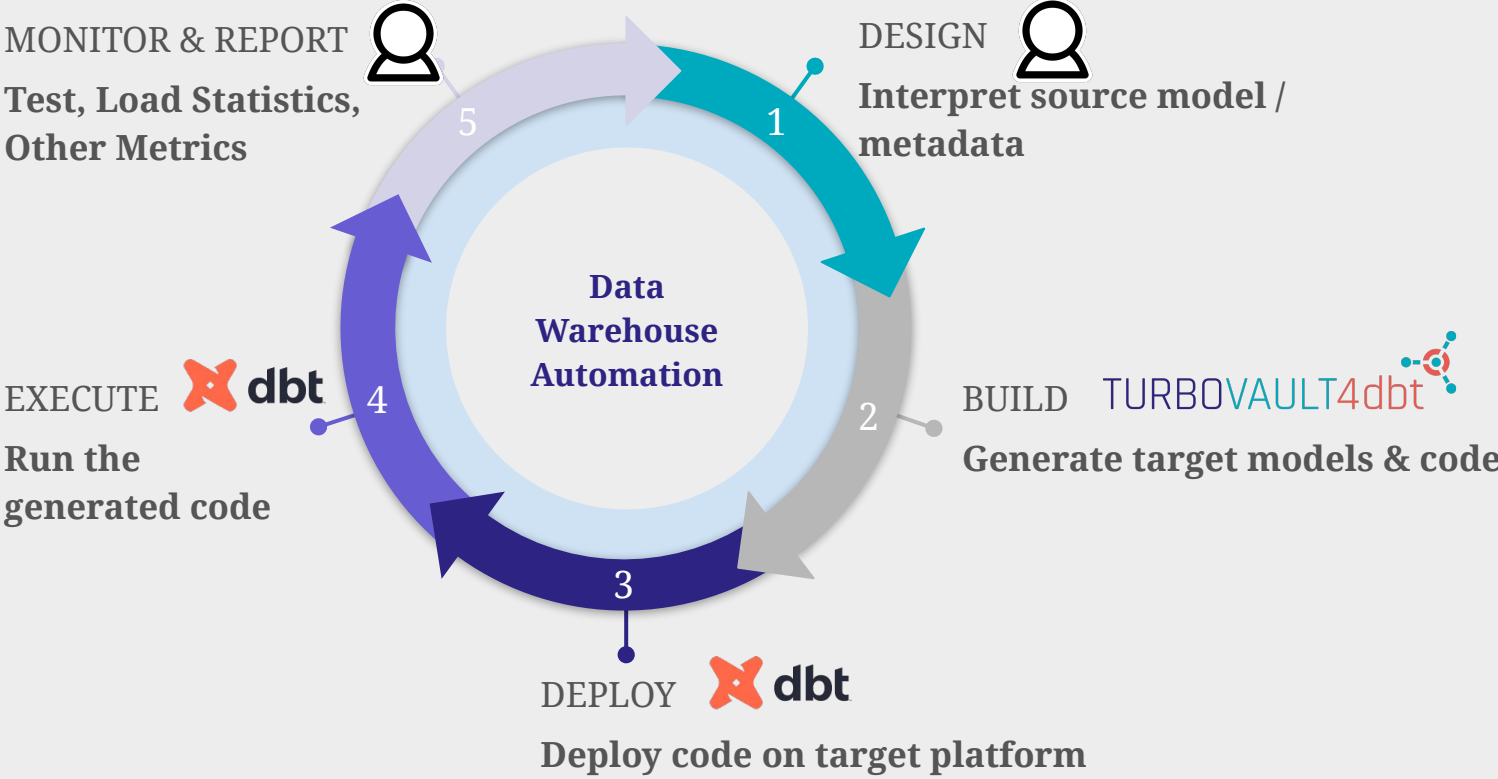
- Open-source tool for automatic dbt model generation
- Utilizes datavault4dbt templates for structured output
- Uses a relational metadata store describing the Data Vault Model
- Python scripts to interpret Metadata
- Generation of datavault4dbt dbt models, Test definitions, and ER diagrams



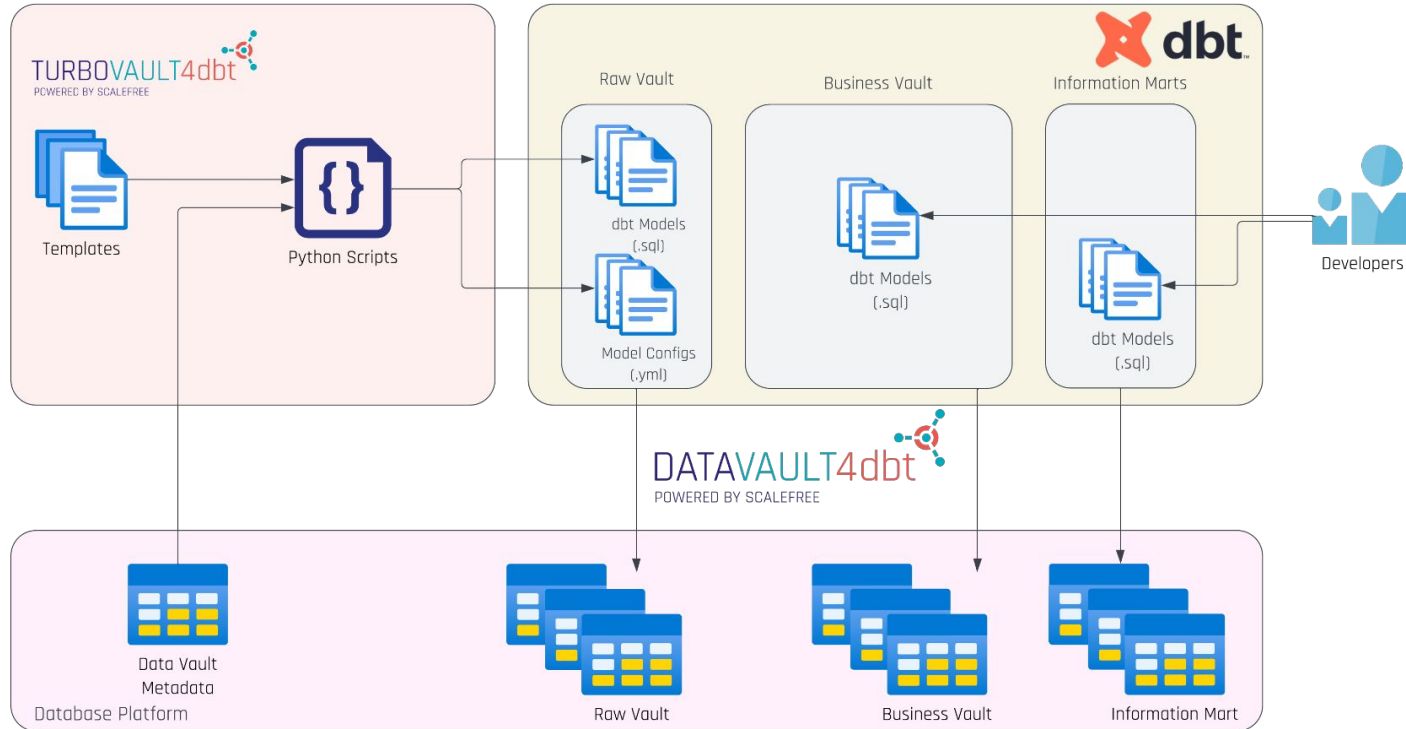
LIFE CYCLE OF DW AUTOMATION (BEFORE)



LIFE CYCLE OF DW AUTOMATION (AFTER)



DEVELOPMENT WORKFLOW



The background is a solid teal color. On the left side, there is a vertical light blue bar. In the center and right, there are several overlapping, semi-transparent geometric shapes: a circle, a horizontal bar, and a larger circular structure with concentric rings and radial segments, resembling a stylized gear or a data visualization element.

DATAVAULT4DBT DEMO

The background is a solid teal color with several abstract, semi-transparent geometric shapes. On the left, there is a vertical teal bar. In the center and right, there are various shapes including circles, rectangles, and larger circular patterns with internal segments, all in lighter shades of teal.

RECOMMENDATIONS

BEST PRACTICES FOR DATA VAULT PROJECTS WITH DBT

- Start with writing models on your own
- Automate model generation as soon as all developers know what they are doing
- Stick to DRY principle
- Implement Naming Conventions & Standards from the beginning
- Test & document your data!
- Don't reinvent the wheel → Use dbt packages!

ADDITIONAL DBT PACKAGES WE USE

- **dbt_utils**

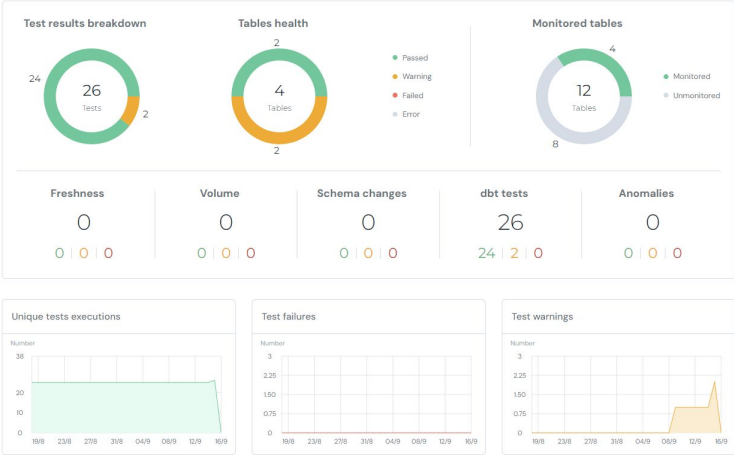
- Collection of helpful macros and utilities that extend the functionality of dbt
- Includes macros for common data modeling tasks, e.g. (un)pivot, star schema helpers or macros for complex SQL logic (like recursive CTEs, window functions, etc.)

- **Elementary**

- Data reliability tool to integrate data observability and quality checks into the dbt workflow
- dbt artifacts uploader, automated monitors, data quality dashboard, anomaly detection tests & more

- **dbt Coverage**

- Coverage reports for docs and tests of your dbt project



DBT AND DATA VAULT 2.0

THE ULTIMATE DATA WAREHOUSING COURSE

Delve into dbt Cloud & benefit from insights shared by our dbt experts. Gain hands-on experience implementing Data Vault 2.0 solutions and explore our package DataVault4dbt!

SCAN & SAVE YOUR SEAT



<https://scalefr.ee/Meb7PY>

DBT-TALK - HUB & SPEAK

EXPERT SESSION

Join us for a monthly expert session: dbt Talk – Hub and Speak

- Send in questions beforehand to be answered by our dbt experts
- Every first thursday of the month

SCAN & SIGN UP FOR FREE



<https://scalefr.ee/dbt-talk>

YOUR DATA VAULT CONTENT PROVIDER

1

KNOWLEDGE
ARTICLES

> 75 **free articles** on
blog

2

MONTHLY
WEBINAR

New topic every
month
> 50 **recordings**

3

DATA VAULT
FRIDAY

Your questions -
every Friday.
> 100 **recordings**

4

EXPERT
SESSIONS

Monthly sessions for
various topics (e.g.
Agile management)

5

PARTNER
HANDS-ON
SESSIONS

Regular Hands-On
sessions. E.g. with
dbt

6

OPEN SOURCE
PACKAGES

e.g. DataVault4dbt

7

DATA VAULT
SOLUTIONS

Various solutions in
our Knowledge hub.

8

CASE STUDIES

>50 **case studies** on
successful
engagements.



SCAN & DISCOVER
OUR KNOWLEDGE
BASE

scalefree.com/knowledge/

ANY MORE QUESTIONS?

QUESTIONS?

THANK YOU FOR YOUR ATTENTION!

What challenges do you have in your project?

SCALEFREE SUPPORTS YOU AND YOUR TEAM WITH KNOW-HOW TRANSFER

- TRAINING
- CONSULTING
- DELIVERY

www.scalefree.com/event-page
www.scalefree.com/#webinars

Feel free to contact us via sales@scalefree.com

- Webinar dates on website www.scalefree.com
- Package available on github.com/ScalefreeCOM/datavault4dbt
- Further information via sales@scalefree.com