Qlik | iNFOTRUST

Reference Architecture of Data Management in Cloud

SIMAS BARANAUSKAS
System Architect, Infotrust

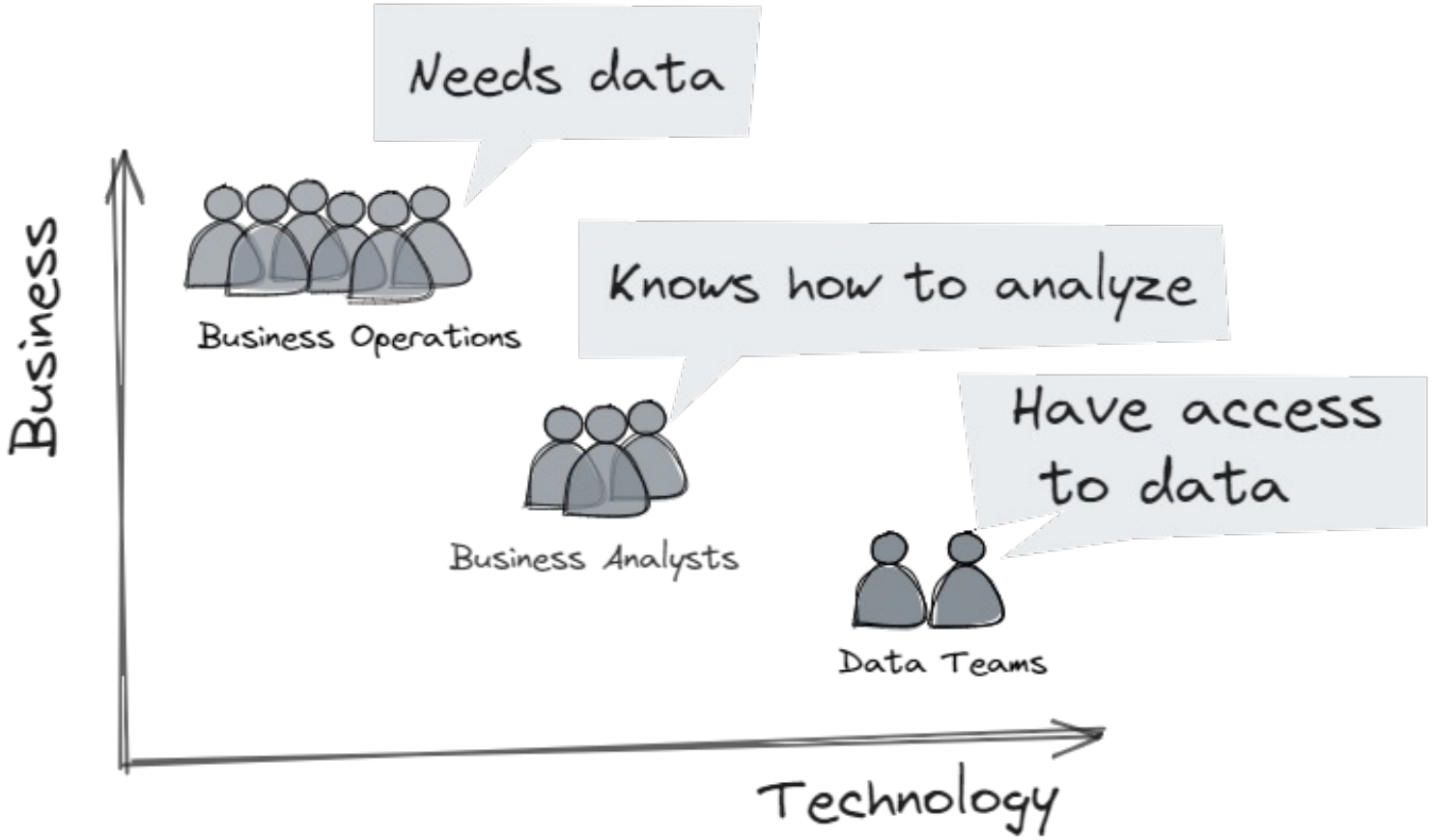
MĀRIS SVILĀNS
Head of Sales, Infotrust

QLIK BALTICS ONLINE #2

QLIK AND SNOWFLAKE: SHAPE YOUR DATA

DATA MANAGEMENT IN THE CLOUD

OR MODERN DATA STACK

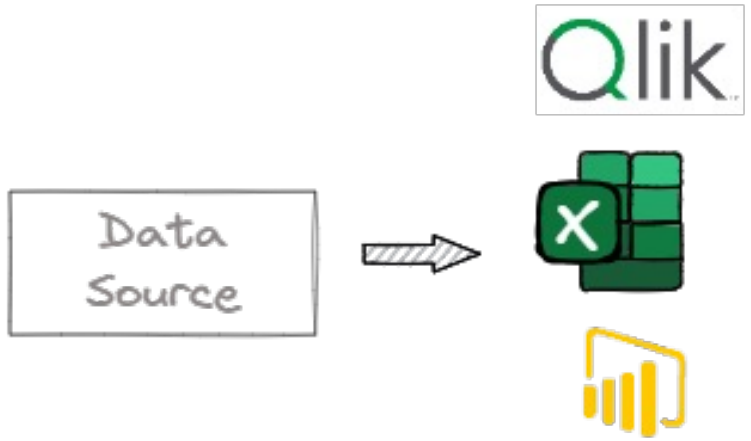


THE STAGES OF DATA ACCESS

1 SOURCE

SOURCE stage is suitable as long as:

- Limited personnel, with only a select few working with data
- Minimal data requirements
- Only a handful of small data sources
- New visuals are created by few technicall people



Transition to DATA LAKE is needed when:

- Data needs to be accessed across multiple locations
- An increasing number of users require data access
- Data processing has performance issues
- Business users require the ability to independently create visualizations

THE STAGES OF DATA ACCESS

2 DATA LAKE

DATA LAKE is still sufficient if:

- A core group is familiar with the diverse data structures
- Limited time or skills for data modeling
- Large datasets require performant queries

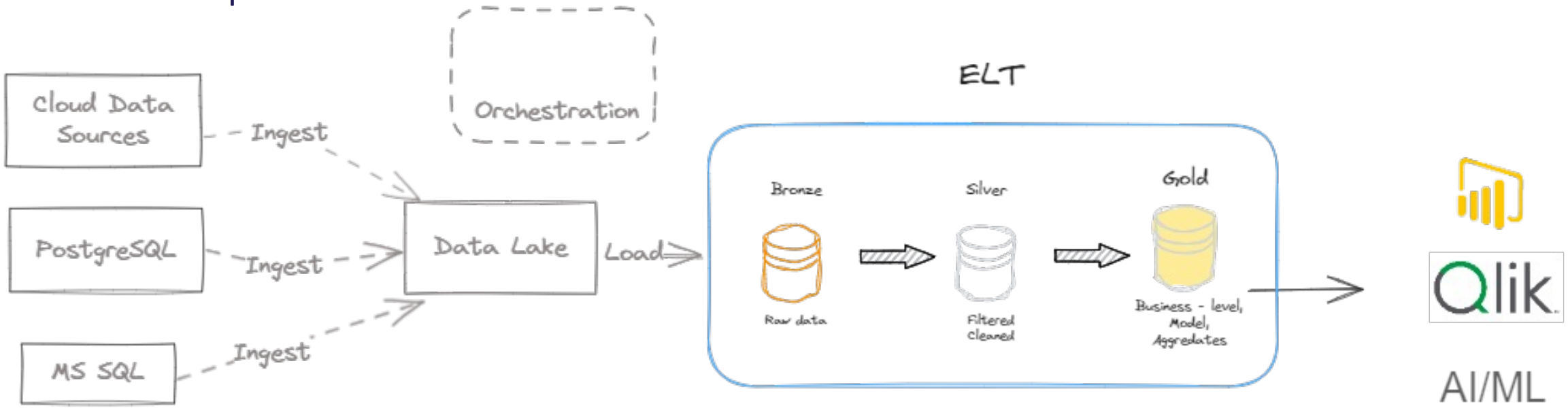
Transition to DATA WAREHOUSE if:

- Many groups of people will need data
- A reliable source of truth is needed or expose via semantic layer
- Solve integrity issues
- Separation of data structure from changing transactional sources is necessary
- AI/ ML use cases start to appear

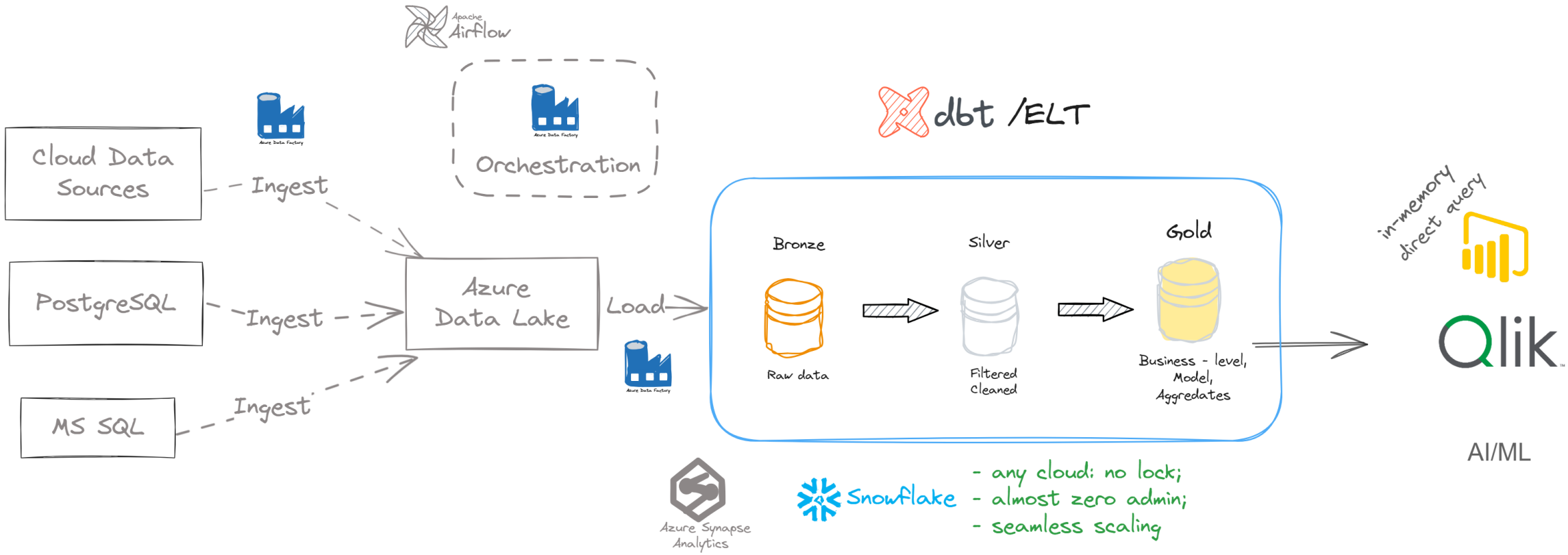
THE STAGES OF DATA ACCESS

3 DATA WAREHOUSE

- It is final stage, in addition DATA MARTS can be created to cater the needs of particular business functions/departments
- The democratization of data now is helping business to explore and understand without external help



REFERENCE ARCHITECTURE



DATA LAKE (DATA STORAGE)

Benefits:

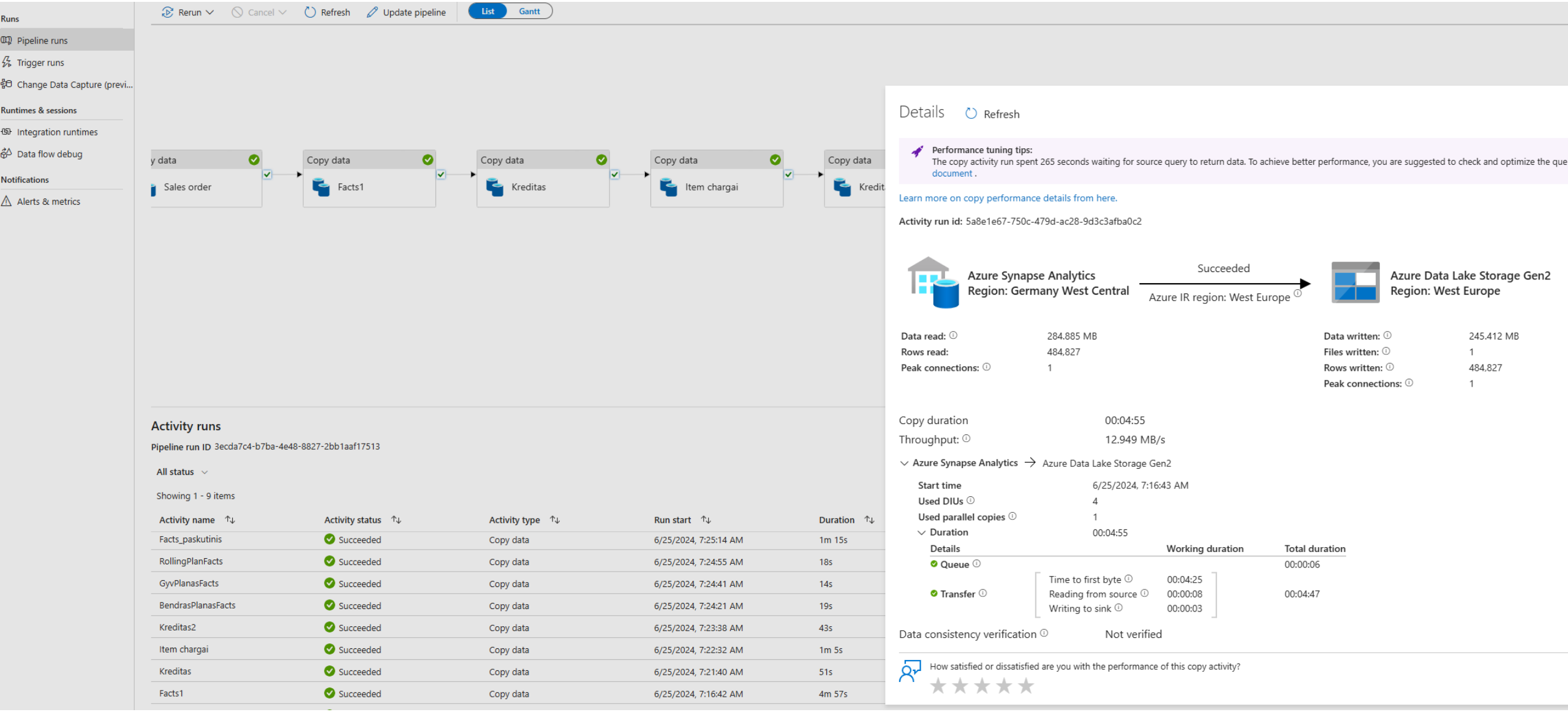
- **Durable and highly available.** Redundancy ensures that your data is safe in the event of transient hardware failures. You can also opt to replicate data across data centers or geographical regions for additional protection from local catastrophe or natural disaster. Data replicated in this way remains highly available in the event of an unexpected outage.
- **Secure.** All data written to an Azure storage account is encrypted by the service. Azure Storage provides you with fine-grained control over who has access to your data.
- **Scalable.** Azure Storage is designed to be massively scalable to meet the data storage and performance needs of today's applications.
- **Managed.** Azure handles hardware maintenance, updates, and critical issues for you.
- **Accessible.** Data in Azure Storage is accessible from anywhere in the world over HTTP or HTTPS. Microsoft provides client libraries for Azure Storage in a variety of languages, including .NET, Java, Node.js, Python, Go, and others, as well as a mature REST API. Azure Storage supports scripting in Azure PowerShell or Azure CLI. And the Azure portal and Azure Storage Explorer offer easy visual solutions for working with your data.

ORCHESTRATION / INGESTION

Benefits:

- **Monitoring**
- Complex (cycling) data loads
- Out of the box connectors
- Integration runtime for on-premise data
- Easy to use and scale (re-use pipelines)

ORCHESTRATION / INGESTION

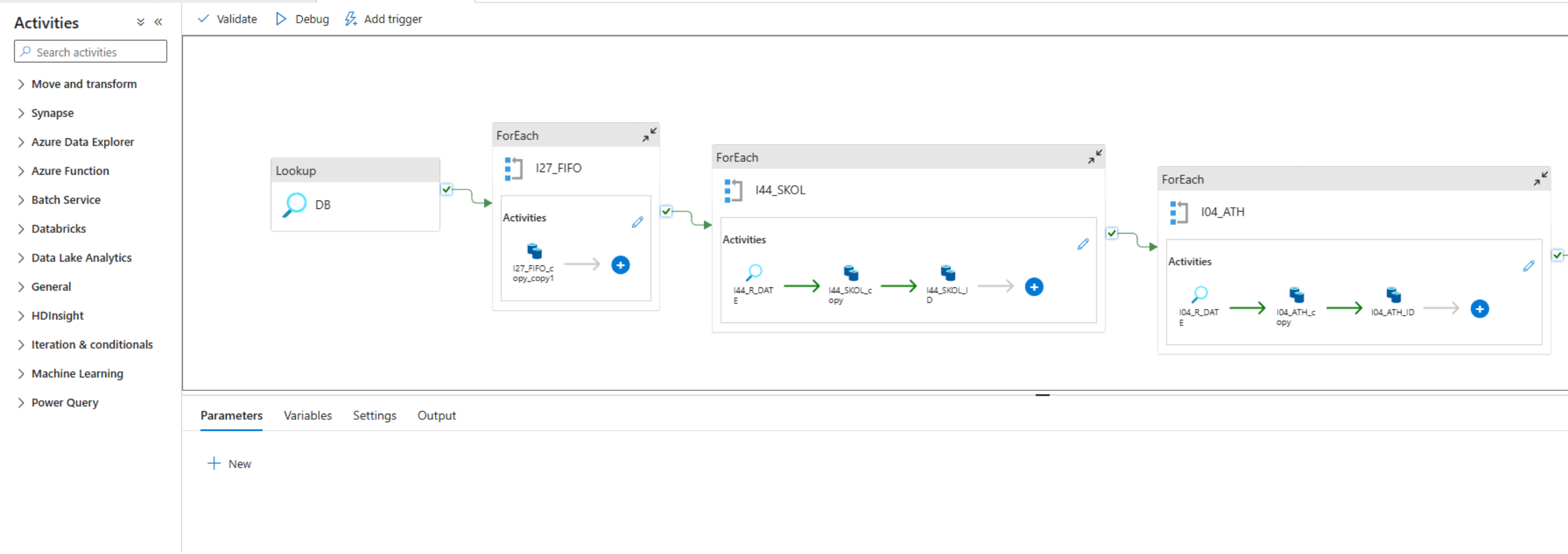


ORCHESTRATION / INGESTION

Benefits:

- Monitoring
- **Complex (cycling) data loads**
- Out of the box connectors
- Integration runtime for on-premise data
- Easy to use and scale (re-use pipelines)

ORCHESTRATION / INGESTION



Copy to clipboard

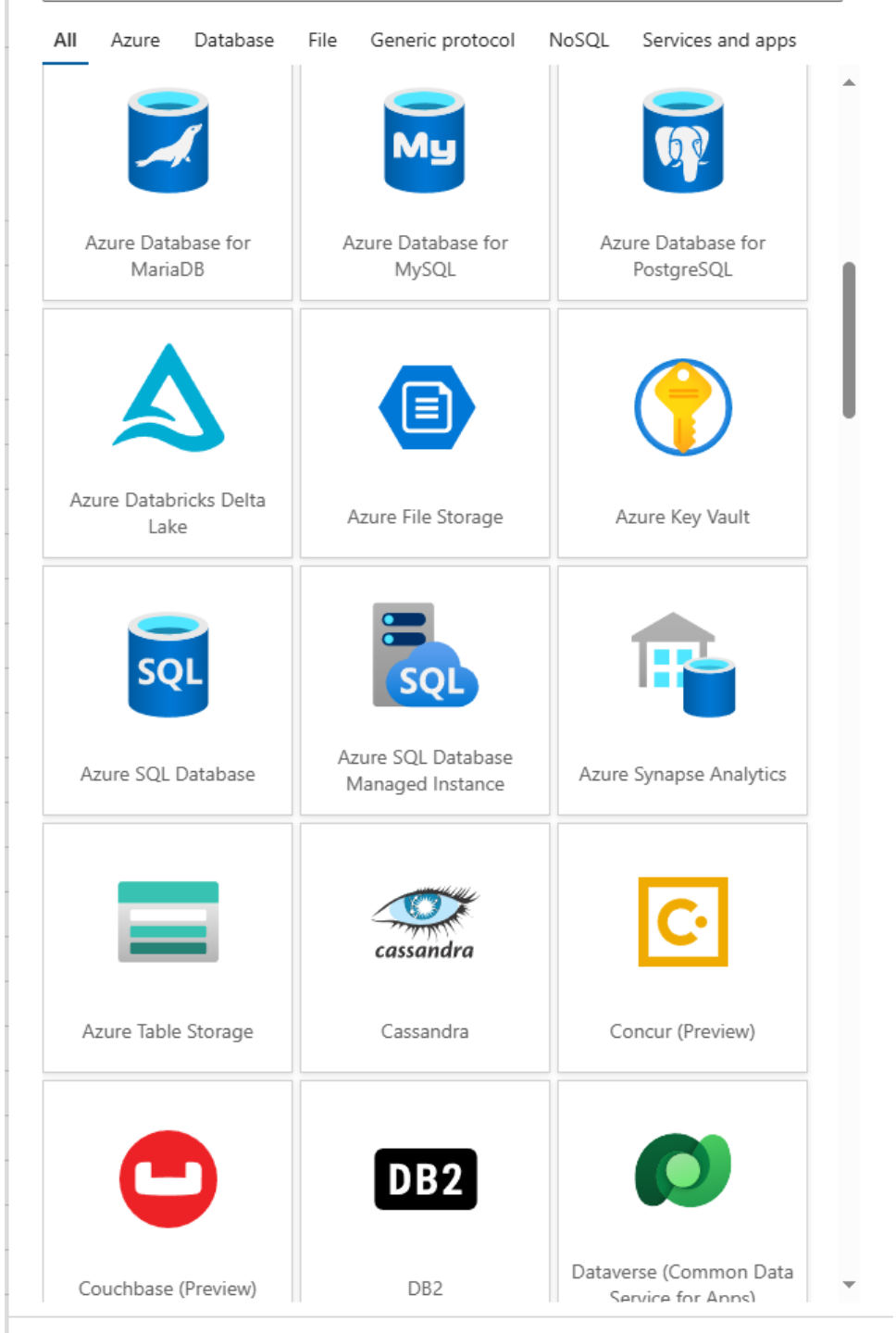
```
        "activities": [
            {
                "name": "I03_DKD",
                "type": "ForEach",
                "dependsOn": [
                    {
                        "activity": "I04_ATH",
                        "dependencyConditions": [
                            "Succeeded"
                        ]
                    }
                ],
                "userProperties": [],
                "typeProperties": {
                    "items": {
                        "value": "@activity('DB').output.value",
                        "type": "Expression"
                    },
                    "isSequential": false,
                    "activities": [
                        {
                            "name": "I03_R_DATE",
                            "description": "",
                            "type": "Lookup",
                            "dependsOn": [],
                            "policy": {
                                "timeout": "7.00:00:00",
                                "retry": 0,
                                "retryIntervalInSeconds": 30,
                                "secureOutput": false,
                                "secureInput": false
                            },
                            "userProperties": [],
                            "typeProperties": {
                                "source": {
                                    "type": "AzureSqlSource",
                                    "sqlReaderQuery": {
                                        "value": "SELECT isnull(max([I03_R_DATE]),'1753-01-01')  as I03_R_DATE\n  FROM [stg].[I03_DKD]\n  where DB = '@{item().DB}'",
                                        "type": "Expression"
                                    },
                                    "queryTimeout": "02:00:00",
                                    "partitionOption": "None"
                                },
```

ORCHESTRATION / INGESTION

Benefits:

- Monitoring
- Complex (cycling) data loads
- **Out of the box connectors**
- Integration runtime for on-premise data
- Easy to use and scale (re-use pipelines)

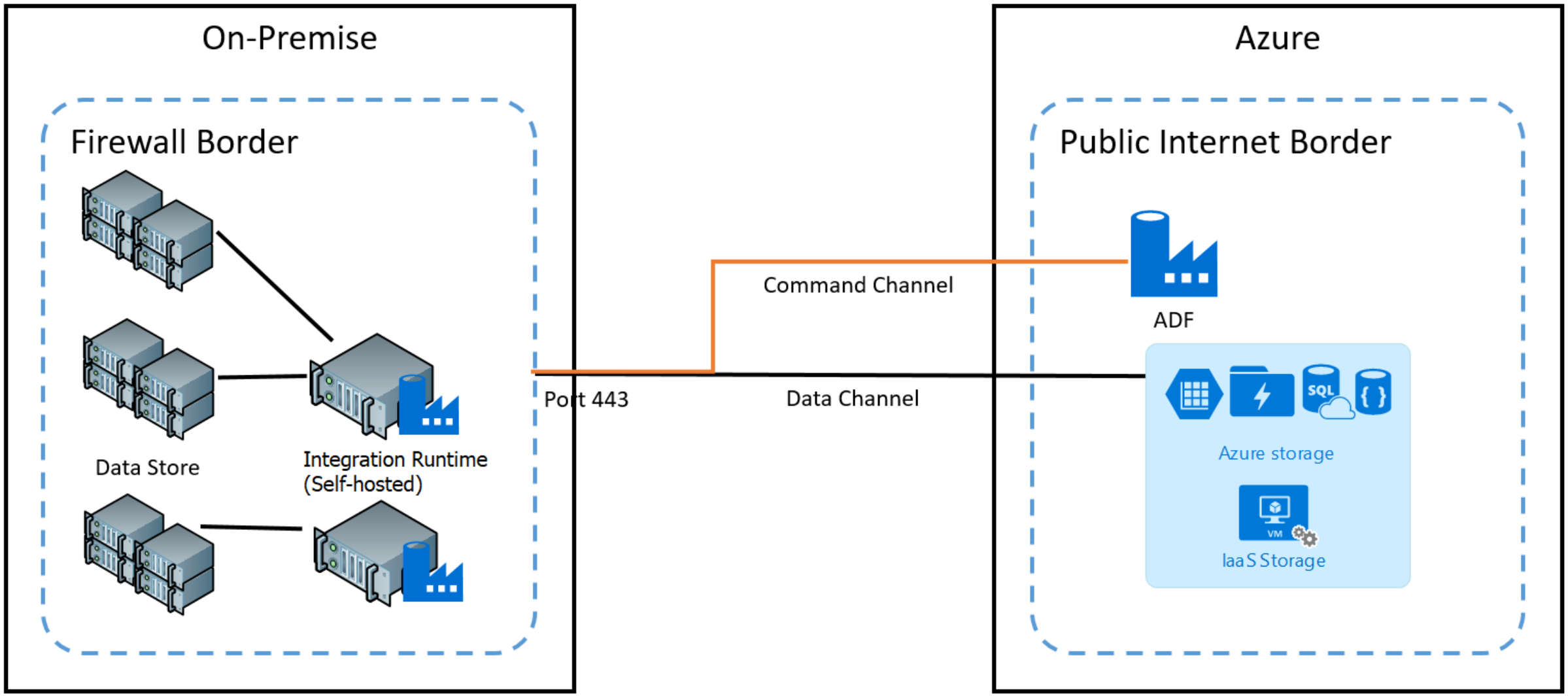
ORCHESTRATION / INGESTION



ORCHESTRATION / INGESTION

Benefits:

- Monitoring
- Complex (cycling) data loads
- Out of the box connectors
- **Integration runtime for on-premise data**
- Easy to use and scale (re-use pipelines)

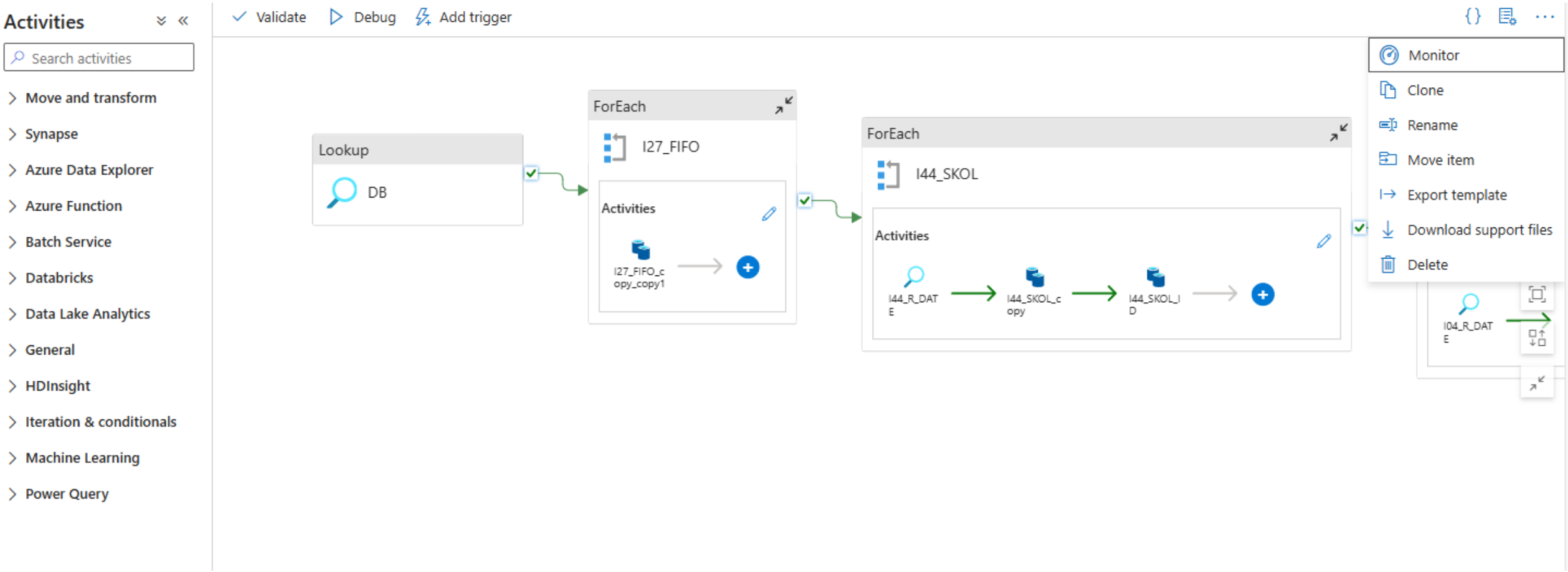

On-Premise
Firewall Border
Data Store
Integration Runtime
(Self-hosted)
Port 443
Command Channel
Data Channel
Azure
Public Internet Border
ADF
SQL
Azure storage
VM
IaaS Storage

ORCHESTRATION / INGESTION

Benefits:

- Monitoring
- Complex (cycling) data loads
- Out of the box connectors
- Integration runtime for on-premise data
- **Easy to use and scale (re-use pipelines)**

ORCHESTRATION / INGESTION



DATA WAREHOUSE -> STORAGE and COMPUTING -> SNOWFLAKE/SYNAPSE/...

Benefits:

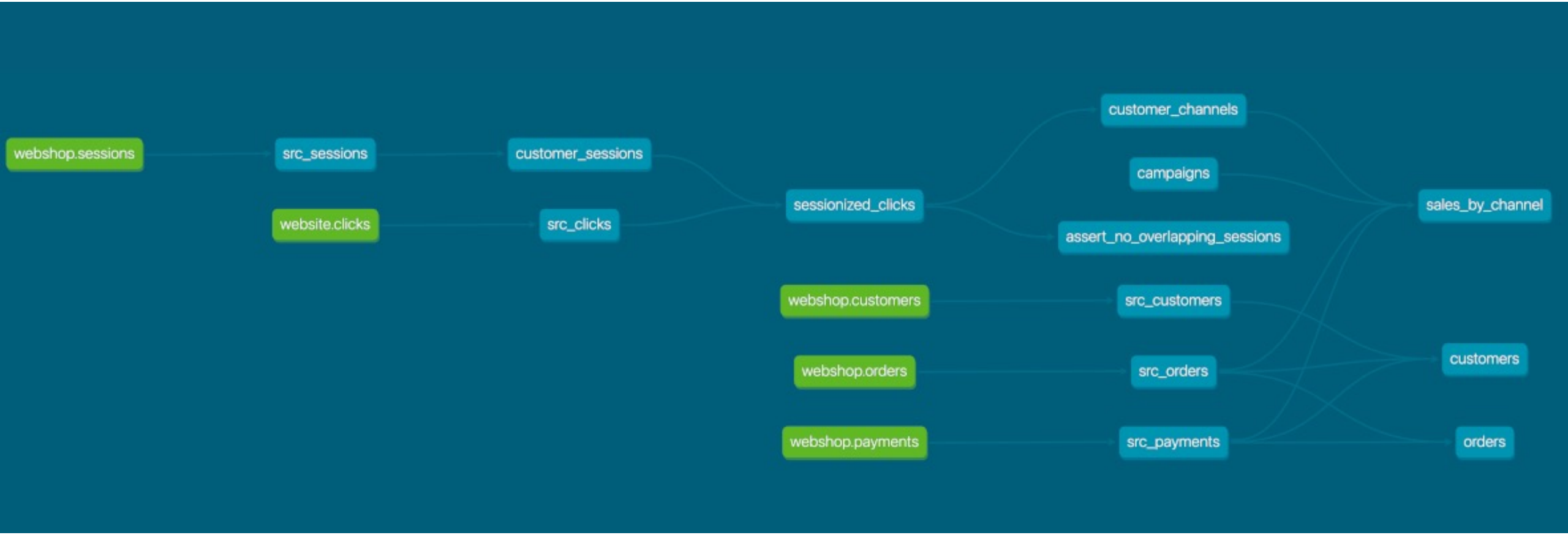
- Performance
- Scalability
- Intermediate results
- Possibility to collect additional data (e.g. snap-shots, that can not be traced later)
- Ability to analyze data with different tools
- Access to data for complete self-service
- Data discovery

DBT

Benefits:

- **Management of complex lineage**
- Automatic documentation of Your solution
- Monitoring of intermediate results, testing
- Version control
- Templating (jinja engine)

DBT

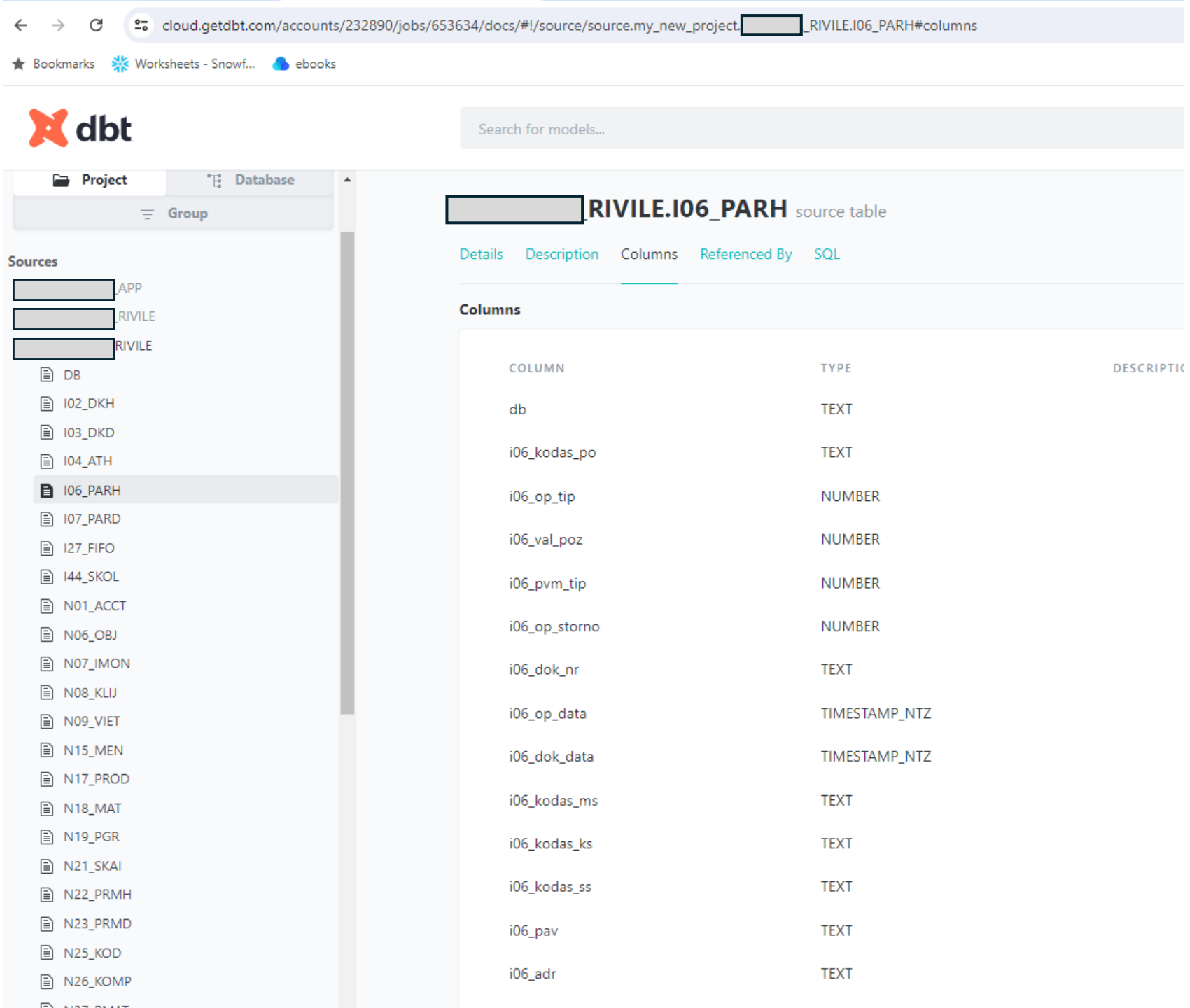

webshop.sessions
src_sessions
customer_sessions
website.clicks
src_clicks
sessionized_clicks
customer_channels
campaigns
assert_no_overlapping_sessions
webshop.customers
src_customers
webshop.orders
src_orders
webshop.payments
src_payments
sales_by_channel
customers
orders

DBT

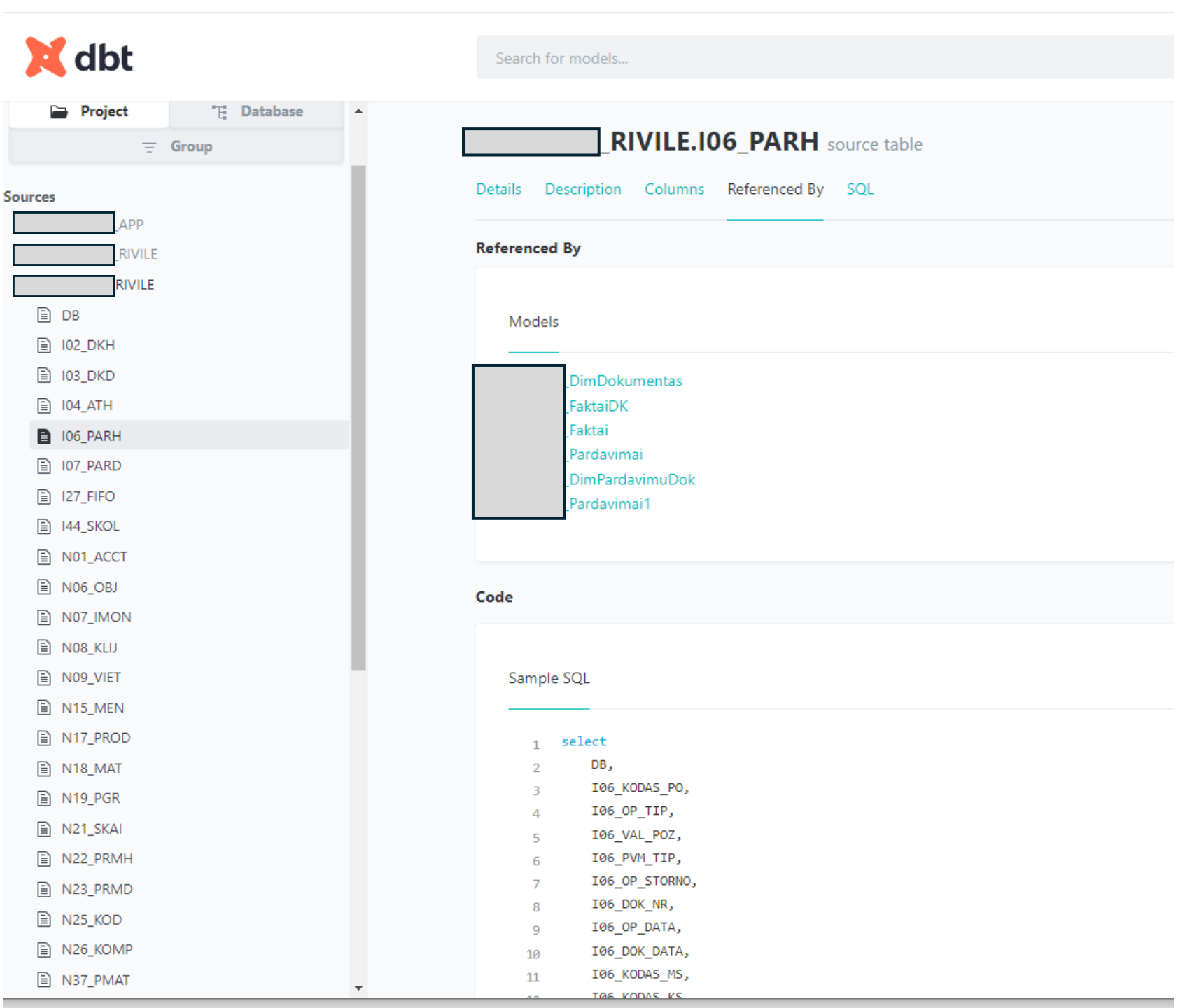
Benefits:

- Management of complex lineage
- **Automatic documentation of Your solution**
- Monitoring of intermediate results, testing
- Version control
- Templating (jinja engine)

DBT

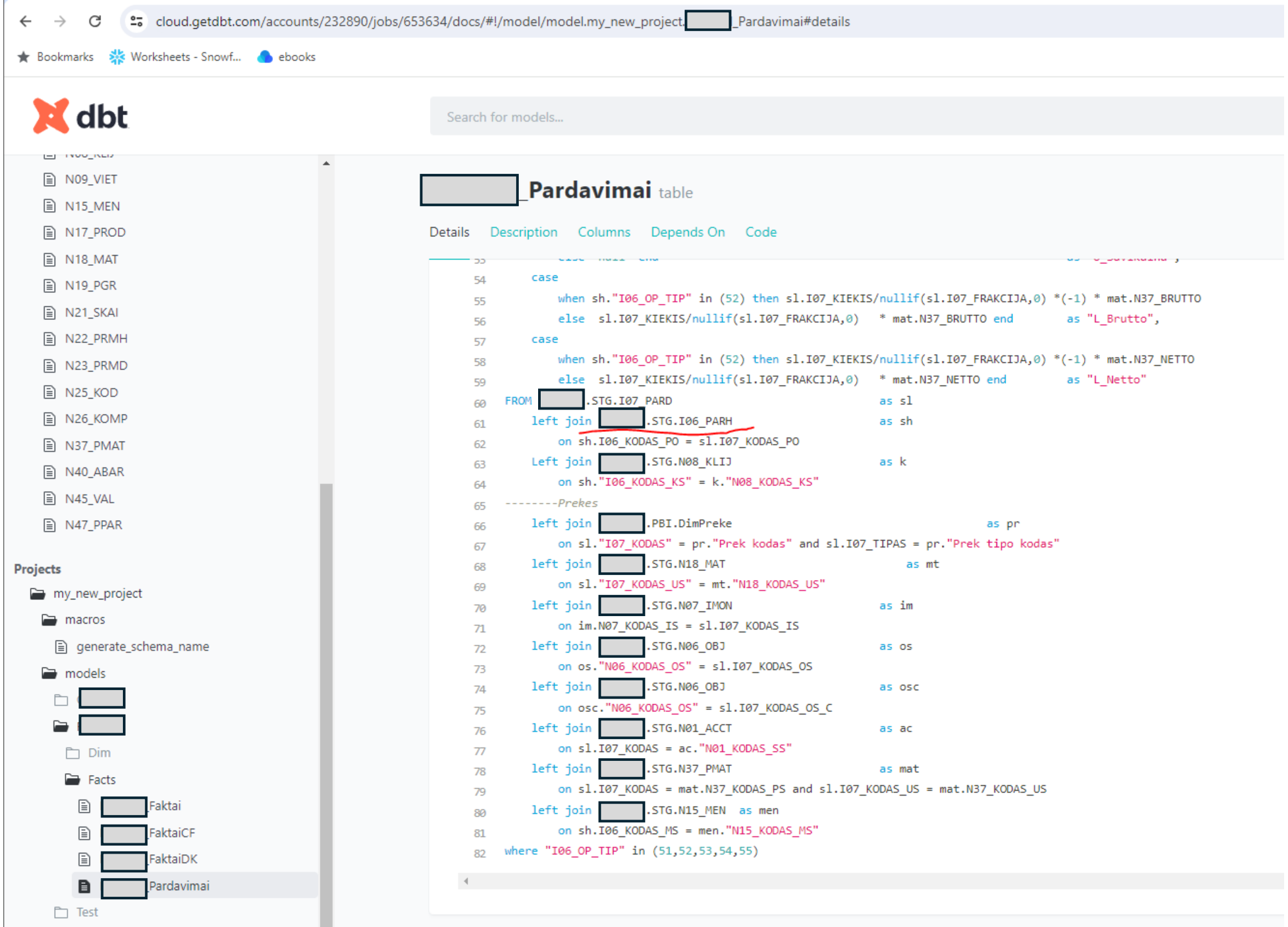

cloud.getdbt.com/accounts/232890/jobs/653634/docs/#!/source/source.my_new_project.[redacted]_RIVILE.I06_PARH#columns
Bookmarks
Worksheets - Snowf...
ebooks
dbt
Search for models...
Project
Database
Group
Sources
[redacted]APP
[redacted]RIVILE
[redacted]RIVILE
DB
I02_DKH
I03_DKD
I04_ATH
I06_PARH
I07_PARD
I27_FIFO
I44_SKOL
N01_ACCT
N06_OBJ
N07_IMON
N08_KLIJ
N09_VIET
N15_MEN
N17_PROD
N18_MAT
N19_PGR
N21_SKAI
N22_PRMH
N23_PRMD
N25_KOD
N26_KOMP
[redacted]_RIVILE.I06_PARH source table
Details
Description
Columns
Referenced By
SQL
Columns
COLUMN
TYPE
DESCRIPTI
db
TEXT
i06_kodas_po
TEXT
i06_op_tip
NUMBER
i06_val_poz
NUMBER
i06_pvm_tip
NUMBER
i06_op_storno
NUMBER
i06_dok_nr
TEXT
i06_op_data
TIMESTAMP_NTZ
i06_dok_data
TIMESTAMP_NTZ
i06_kodas_ms
TEXT
i06_kodas_ks
TEXT
i06_kodas_ss
TEXT
i06_pav
TEXT
i06_adr
TEXT

DBT


dbt
Search for models...
Project
Database
Group
Sources
[redacted]_APP
[redacted]_RIVILE
[redacted]RIVILE
DB
I02_DKH
I03_DKD
I04_ATH
I06_PARH
I07_PARD
I27_FIFO
I44_SKOL
N01_ACCT
N06_OBJ
N07_IMON
N08_KLIJ
N09_VIET
N15_MEN
N17_PROD
N18_MAT
N19_PGR
N21_SKAI
N22_PRMH
N23_PRMD
N25_KOD
N26_KOMP
N37_PMAT
[redacted]_RIVILE.I06_PARH source table
Details
Description
Columns
Referenced By
SQL
Referenced By
Models
[redacted]_DimDokumentas
[redacted]_FaktaiDK
[redacted]_Faktai
[redacted]_Pardavimai
[redacted]_DimPardavimuDok
[redacted]_Pardavimai1
Code
Sample SQL
1 select
2 DB,
3 I06_KODAS_PO,
4 I06_OP_TIP,
5 I06_VAL_POZ,
6 I06_PVM_TIP,
7 I06_OP_STORNO,
8 I06_DOK_NR,
9 I06_OP_DATA,
10 I06_DOK_DATA,
11 I06_KODAS_MS,

DBT

-


cloud.getdbt.com/accounts/232890/jobs/653634/docs/#!/model/model.my_new_project.[redacted]_Pardavimai#details
Bookmarks
Worksheets - Snowf...
ebooks
dbt
Search for models...
N09_VIET
N15_MEN
N17_PROD
N18_MAT
N19_PGR
N21_SKAI
N22_PRMH
N23_PRMD
N25_KOD
N26_KOMP
N37_PMAT
N40_ABAR
N45_VAL
N47_PPAR
Projects
my_new_project
macros
generate_schema_name
models
[redacted]
[redacted]
Dim
Facts
[redacted]Faktai
[redacted]FaktaiCF
[redacted]FaktaiDK
[redacted]Pardavimai
Test
[redacted]_Pardavimai table
Details
Description
Columns
Depends On
Code
54 case
55 when sh."I06_OP_TIP" in (52) then sl.I07_KIEKIS/nullif(sl.I07_FRAKCIJA,0) *(-1) * mat.N37_BRUTTO
56 else sl.I07_KIEKIS/nullif(sl.I07_FRAKCIJA,0) * mat.N37_BRUTTO end as "L_Brutto",
57 case
58 when sh."I06_OP_TIP" in (52) then sl.I07_KIEKIS/nullif(sl.I07_FRAKCIJA,0) *(-1) * mat.N37_NETTO
59 else sl.I07_KIEKIS/nullif(sl.I07_FRAKCIJA,0) * mat.N37_NETTO end as "L_Netto"
60 FROM [redacted].STG.I07_PARD as sl
61 left join [redacted].STG.I06_PARH as sh
62 on sh.I06_KODAS_PO = sl.I07_KODAS_PO
63 Left join [redacted].STG.N08_KLIJ as k
64 on sh."I06_KODAS_KS" = k."N08_KODAS_KS"
65 --------Prekes
66 left join [redacted].PBI.DimPreke as pr
67 on sl."I07_KODAS" = pr."Prek kodas" and sl.I07_TIPAS = pr."Prek tipo kodas"
68 left join [redacted].STG.N18_MAT as mt
69 on sl."I07_KODAS_US" = mt."N18_KODAS_US"
70 left join [redacted].STG.N07_IMON as im
71 on im.N07_KODAS_IS = sl.I07_KODAS_IS
72 left join [redacted].STG.N06_OBJ as os
73 on os."N06_KODAS_OS" = sl.I07_KODAS_OS
74 left join [redacted].STG.N06_OBJ as osc
75 on osc."N06_KODAS_OS" = sl.I07_KODAS_OS_C
76 left join [redacted].STG.N01_ACCT as ac
77 on sl.I07_KODAS = ac."N01_KODAS_SS"
78 left join [redacted].STG.N37_PMAT as mat
79 on sl.I07_KODAS = mat.N37_KODAS_PS and sl.I07_KODAS_US = mat.N37_KODAS_US
80 left join [redacted].STG.N15_MEN as men
81 on sh.I06_KODAS_MS = men."N15_KODAS_MS"
82 where "I06_OP_TIP" in (51,52,53,54,55)

DBT

Benefits:

- Management of complex lineage
- Automatic documentation of Your solution
- **Monitoring of intermediate results, testing**
- Version control
- Templating (jinja engine)

DBT

```yaml
version: 2

models:
  - name: int_fx_rates
    description: "An intermediate model that filters stg_knoema_fx_rates"
    columns:
      - name: currency||exchange_date
        tests:
          - unique
          - not_null

  - name: int_unioned_book
    description: "An intermediate model unions the manual_book csvs"
    columns:
      - name: instrument
        tests:
          - not_null
          - relationships:
              to: ref('int_knoema_stock_history')
              field: company_symbol

  - name: int_knoema_stock_history
    description: "An intermediate model that pivots the stg_knoema_stock_history model b
    columns:
      - name: company_symbol||stock_date
        tests:
          - not_null
          - unique
```

DBT

Data Tests Overview

Status	Pass	Fail	Error	Last Snapshot
Warning: some data quality issues were detected	6	1	0	23/04/2022 8:59:29 AM

Test	Quality Tag	Status	Database	Table/View	Column
dbt_expectations_expect_table_column_count_to_be_between_my_table1_4__1	unkonwn	fail	my_database	my_table1	n/a
not_null_my_table1_col1	consistency	pass	my_database	my_table1	col1
unique_my_table1_col1	consistency	pass	my_database	my_table1	col1
not_null_my_table2_col1	unkonwn	pass	my_database	my_table2	col1
unique_my_table2_col1	accuracy	pass	my_database	my_table2	col1
not_null_my_table1_col2	timeliness	pass	my_database	my_table1	col2
not_null_my_table2_col2	completeness	pass	my_database	my_table2	col2

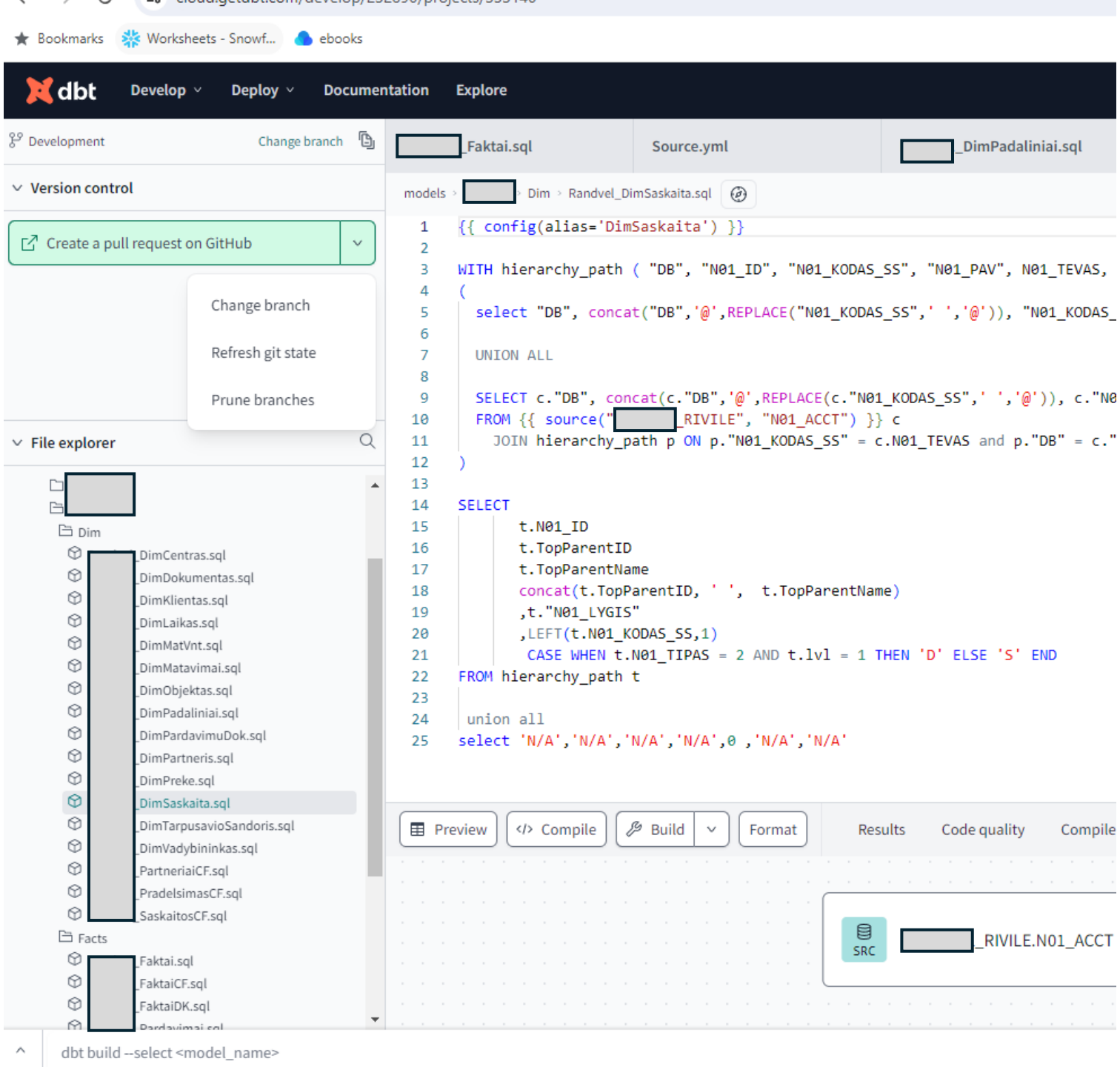
DBT

Benefits:

- Management of complex lineage
- Automatic documentation of Your solution
- Monitoring of intermediate results, testing
- **Version control**
- Templating (jinja engine)

DBT

-


Bookmarks
Worksheets - Snowf...
ebooks
dbt
Develop
Deploy
Documentation
Explore
Development
Change branch
Version control
Create a pull request on GitHub
Change branch
Refresh git state
Prune branches
File explorer
Dim
DimCentras.sql
DimDokumentas.sql
DimKlientas.sql
DimLaikas.sql
DimMatVnt.sql
DimMatavimai.sql
DimObjektas.sql
DimPadaliniai.sql
DimPardavimuDok.sql
DimPartneris.sql
DimPreke.sql
DimSaskaita.sql
DimTarpusavioSandoris.sql
DimVadybininkas.sql
PartneriaiCF.sql
PradelsimasCF.sql
SaskaitosCF.sql
Facts
Faktai.sql
FaktaiCF.sql
FaktaiDK.sql
Faktai.sql
Source.yml
_DimPadaliniai.sql
models
Dim
Randvel_DimSaskaita.sql
{{ config(alias='DimSaskaita') }}
WITH hierarchy_path ("DB", "N01_ID", "N01_KODAS_SS", "N01_PAV", N01_TEVAS,
(
select "DB", concat("DB",'@',REPLACE("N01_KODAS_SS",' ','@')), "N01_KODAS_
UNION ALL
SELECT c."DB", concat(c."DB",'@',REPLACE(c."N01_KODAS_SS",' ','@')), c."N0
FROM {{ source("
_RIVILE", "N01_ACCT") }} c
JOIN hierarchy_path p ON p."N01_KODAS_SS" = c.N01_TEVAS and p."DB" = c."
)
SELECT
t.N01_ID
t.TopParentID
t.TopParentName
concat(t.TopParentID, ' ', t.TopParentName)
,t."N01_LYGIS"
,LEFT(t.N01_KODAS_SS,1)
CASE WHEN t.N01_TIPAS = 2 AND t.lvl = 1 THEN 'D' ELSE 'S' END
FROM hierarchy_path t
union all
select 'N/A','N/A','N/A','N/A',0 ,'N/A','N/A'
Preview
Compile
Build
Format
Results
Code quality
Compile
SRC
_RIVILE.N01_ACCT
dbt build --select <model_name>

Qlik | iNFOTRUST

Thank You!

Simas Baranauskas, System Architect, Infotrust
s.baranauskas@theinfotrust.com
Maris Svilans, Head of Sales, Infotrust
m.svilans@theinfotrust.com

Book a meeting:

QLIK BALTICS ONLINE #2

QLIK AND SNOWFLAKE: SHAPE YOUR DATA