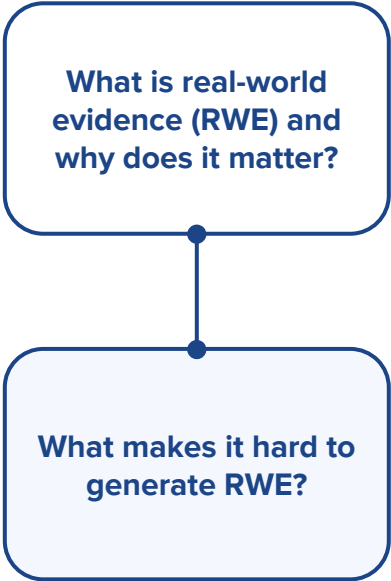


Accelerating privacy-enhancing data processing

Florian Stefan

Staff Software Engineer @ Flatiron Health

**What is real-world
evidence (RWE) and
why does it matter?**



```
graph TD; A[What is real-world evidence (RWE) and why does it matter?] --- B[What makes it hard to generate RWE?];
```

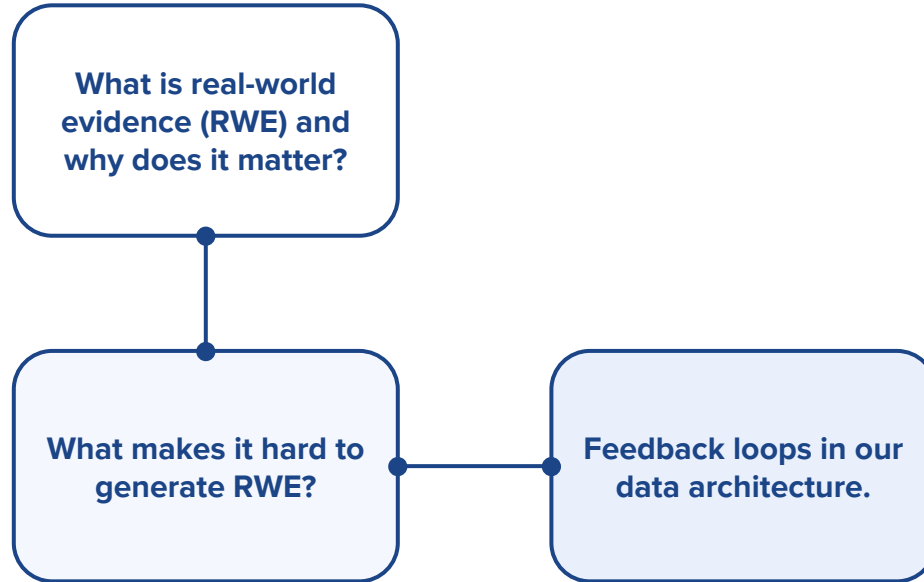
What is real-world evidence (RWE) and why does it matter?

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

What makes it hard to generate RWE?

Feedback loops in our data architecture.

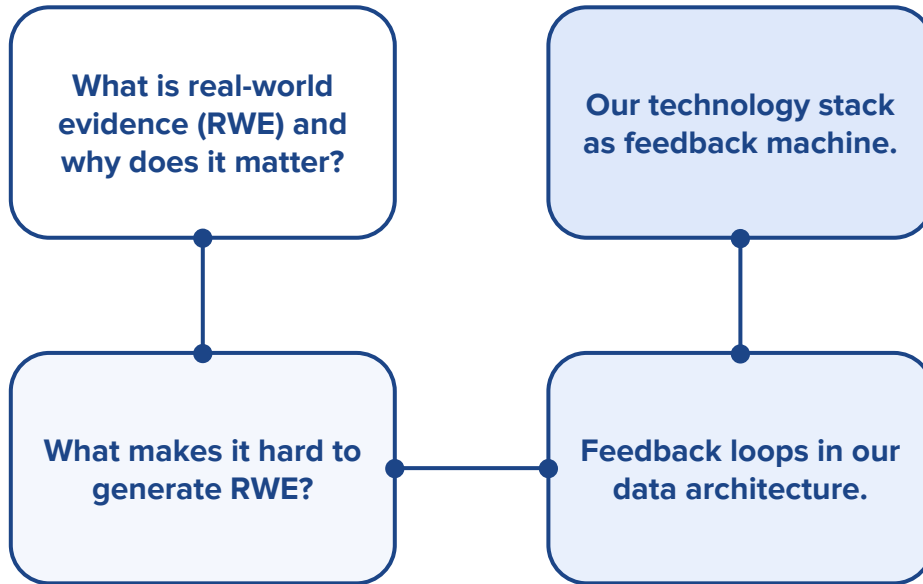


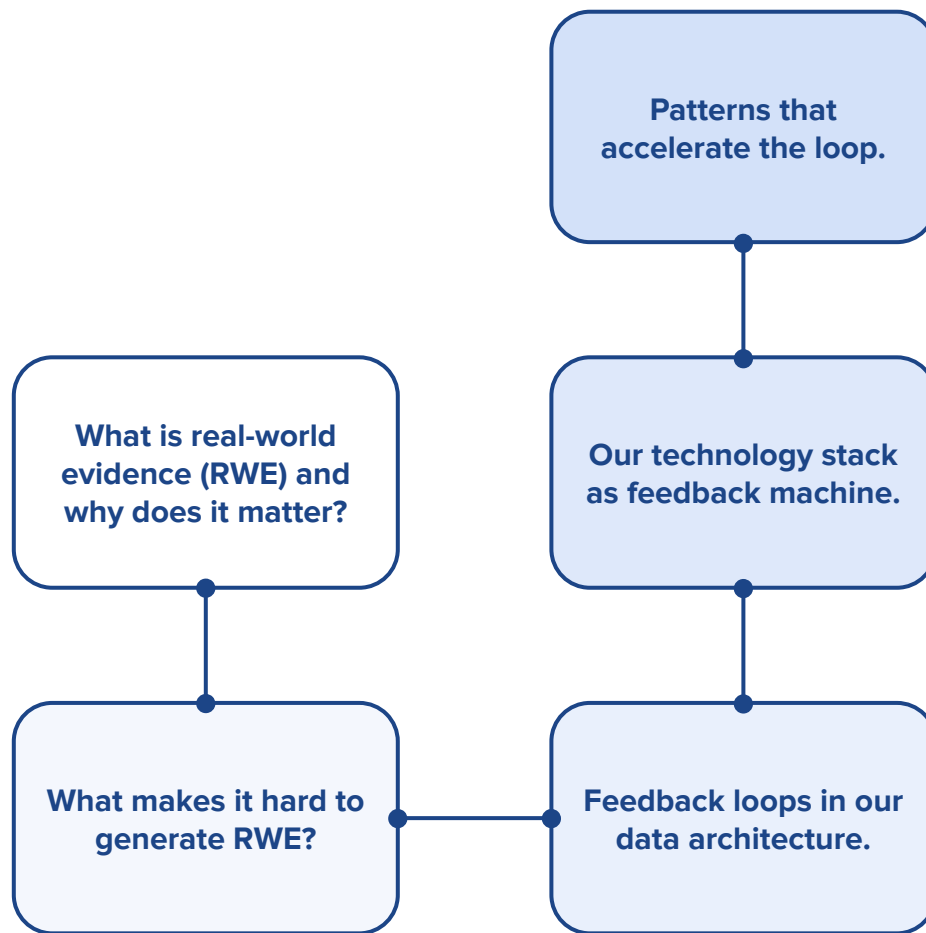
What is real-world evidence (RWE) and why does it matter?

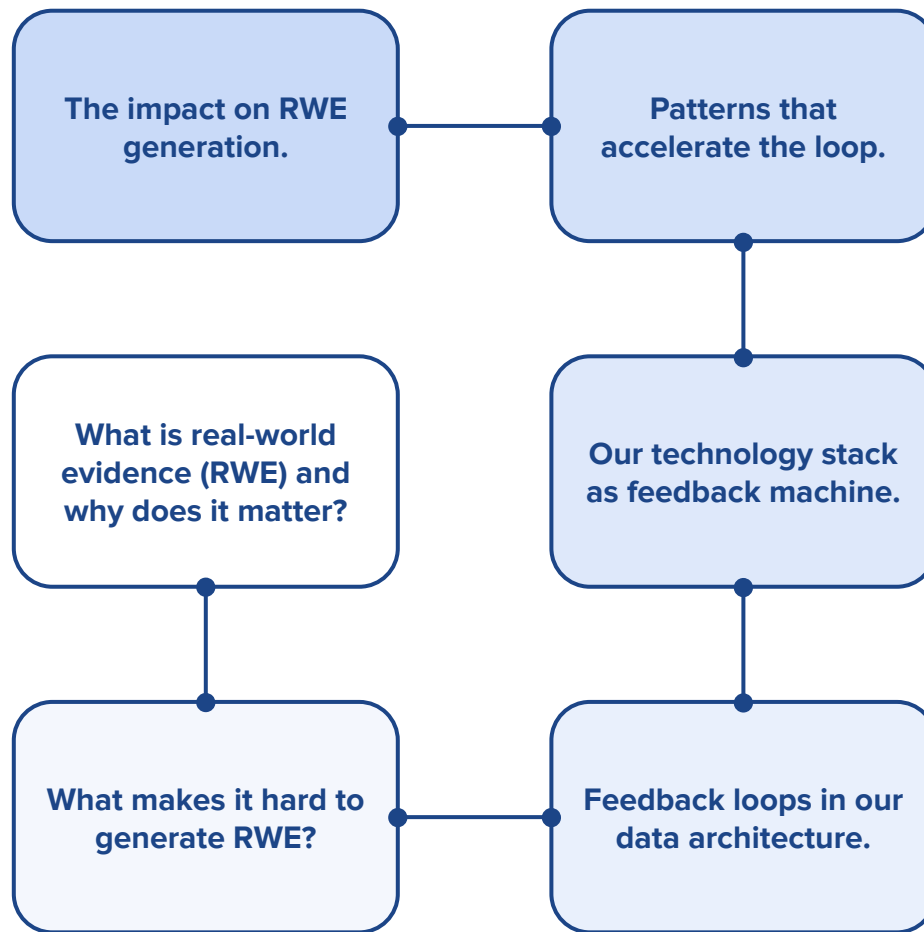
Our technology stack as feedback machine.

What makes it hard to generate RWE?

Feedback loops in our data architecture.







The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

What is real-world evidence (RWE) and why does it matter?

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Generated from clinical data collected during routine care.

- ☐ Data sources include EHRs, laboratory systems, and cancer registries.
- ☐ Combines structured and unstructured data across the healthcare ecosystem.

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Generated from clinical data collected during routine care.

- Data sources include EHRs, laboratory systems, and cancer registries.
- Combines structured and unstructured data across the healthcare ecosystem.

Supports research, regulatory, and clinical decision-making.

- Enables faster and more cost-effective evidence generation.
- Captures real-world treatment outcomes ...
- ... across diverse populations, therapies, and care settings.

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

What is real-world evidence (RWE) and why does it matter?

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

What is real-world evidence (RWE) and why does it matter?

What makes it hard to generate RWE?

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Fragmented and heterogeneous data landscape.

- Data is siloed across institutions, formats, and systems.
- Lack of standards for structuring clinical observations.

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Fragmented and heterogeneous data landscape.

- ❑ Data is siloed across institutions, formats, and systems.
- ❑ Lack of standards for structuring clinical observations.

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

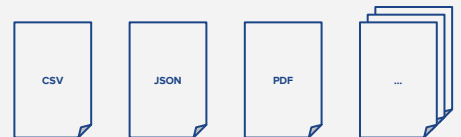
What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

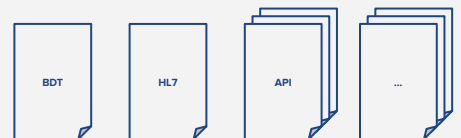
Fragmented and heterogeneous data landscape.

- ❑ Data is siloed across institutions, formats, and systems.
- ❑ Lack of standards for structuring clinical observations.

Data representations



Data exchange formats



Data archive & backup formats



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Fragmented and heterogeneous data landscape.

- Data is siloed across institutions, formats, and systems.
- Lack of standards for structuring clinical observations.

Unstructured, incomplete, and noisy data.

- Critical information often buried in free text, scans, or PDFs.
- Requires OCR, NLP, and manual abstraction to extract meaning.

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Fragmented and heterogeneous data landscape.

- ❑ Data is siloed across institutions, formats, and systems.
- ❑ Lack of standards for structuring clinical observations.

Unstructured, incomplete, and noisy data.

- ❑ Critical information often buried in free text, scans, or PDFs.
- ❑ Requires OCR, NLP, and manual abstraction to extract meaning.



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Fragmented and heterogeneous data landscape.

- Data is siloed across institutions, formats, and systems.
- Lack of standards for structuring clinical observations.

Unstructured, incomplete, and noisy data.

- Critical information often buried in free text, scans, or PDFs.
- Requires OCR, NLP, and manual abstraction to extract meaning.



Patient: Maria T. (DOB: 1966-04-23, MRN: 872394)
Hx: mCRC, T3N2M1, KRAS-mut, DM2, HTN

Pt seen today with daughter for f/u. Reports stable appetite, mild fatigue. Completed 8 cycles FOLFOX + Bev - PD on 8/30/2023 CT (incr. liver mets, new pulm nodule). Plan: switch to FOLFIRI + Bev, pending ANC >1.0 + plt >100. Labs drawn today; port to be placed Thursday (9/21).

Avoid steroids due to poorly controlled DM2 (last A1c: 9.3%). Will start antiemetics (ondansetron PRN; consider olanzapine). Imaging planned for 11/13/2023 to assess response.

Discussed goals of care. Pt wants to continue active tx, prefers to be treated at clinic near home in Trenton if possible.

Note dictated by Dr. Klein, 9/18/2023 at 17:42.

Test	Value	Units
WBC	5.6	$\times 10^9/L$
ANC	0.8	$\times 10^9/L$
HbA1c (from 09/01)	9.3	%

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Fragmented and heterogeneous data landscape.

- ❑ Data is siloed across institutions, formats, and systems.
- ❑ Lack of standards for structuring clinical observations.

Unstructured, incomplete, and noisy data.

- ❑ Critical information often buried in free text, scans, or PDFs.
- ❑ Requires OCR, NLP, and manual abstraction to extract meaning.

Strict regulatory and privacy constraints.

- ❑ Patient data must be handled with strong privacy protections.
- ❑ Access and re-use often restricted across organizations.

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Fragmented and heterogeneous data landscape.

- ❑ Data is siloed across institutions, formats, and systems.
- ❑ Lack of standards for structuring clinical observations.

Unstructured, incomplete, and noisy data.

- ❑ Critical information often buried in free text, scans, or PDFs.
- ❑ Requires OCR, NLP, and manual abstraction to extract meaning.

Strict regulatory and privacy constraints.

- ❑ Patient data must be handled with strong privacy protections.
- ❑ Access and reuse often restricted across organizations.



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Programmed and heterogeneous data collection	Unstructured, incomplete, and noisy data
<ul style="list-style-type: none">• Extracts related patient information, history, and behavior	<ul style="list-style-type: none">• Unstructured information often located in free text, notes, or PDFs
<ul style="list-style-type: none">• Lack of standards for structuring patient information	<ul style="list-style-type: none">• Requires NLP, ML, and manual annotation to process meaning

Search engines	Index
<ul style="list-style-type: none">• Extracts and structures information, creating a searchable index	
<ul style="list-style-type: none">• Searches and retrieves information	



The impact on RWE generation.

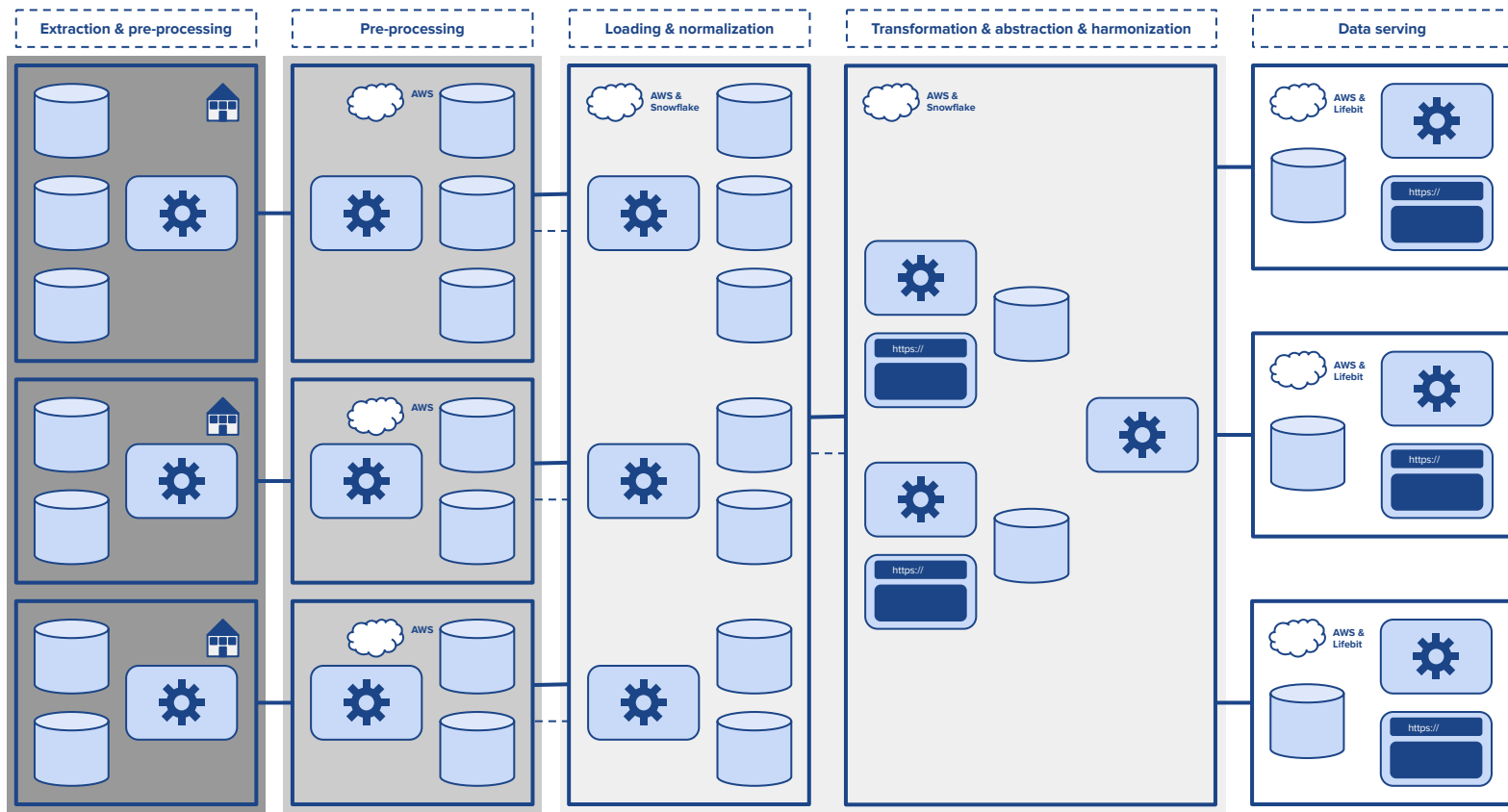
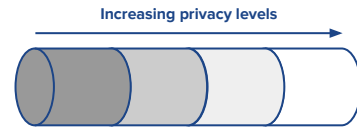
Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

What is real-world evidence (RWE) and why does it matter?

What makes it hard to generate RWE?

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

What is real-world evidence (RWE) and why does it matter?

What makes it hard to generate RWE?

Feedback loops in our data architecture.

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



Clinicians



Data scientists

The impact on RWE generation.

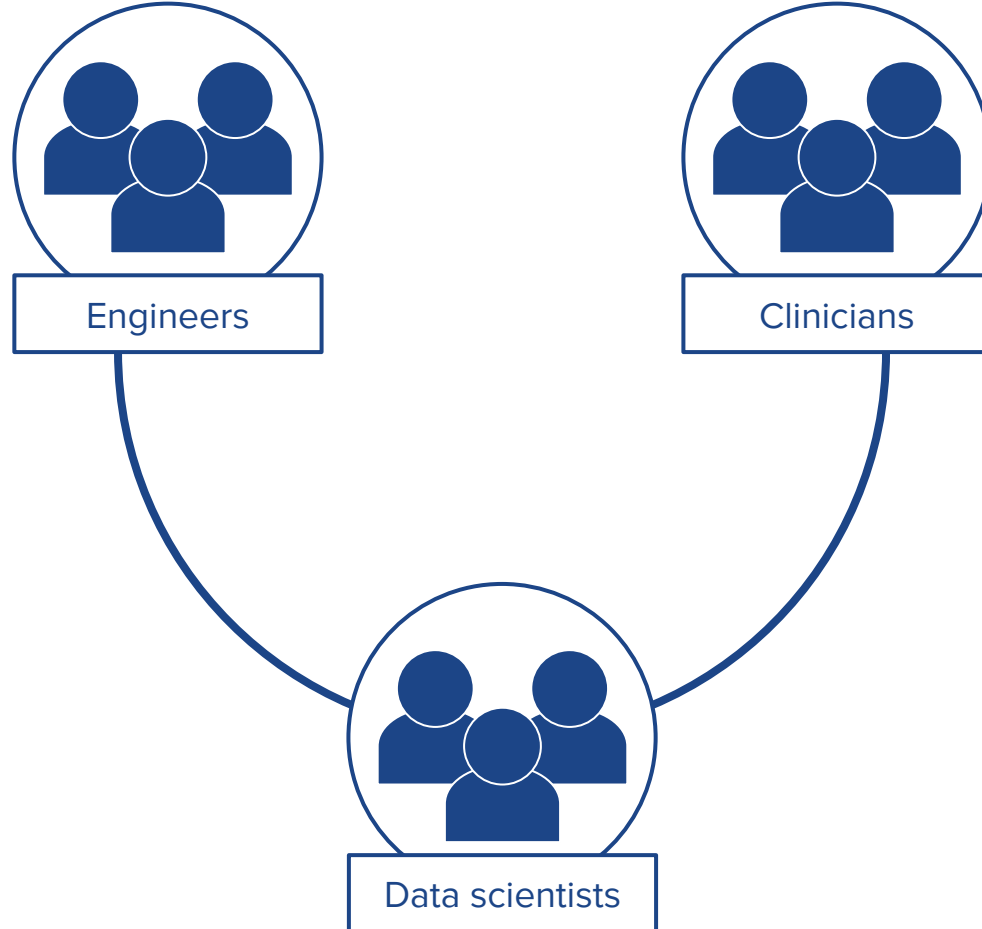
Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

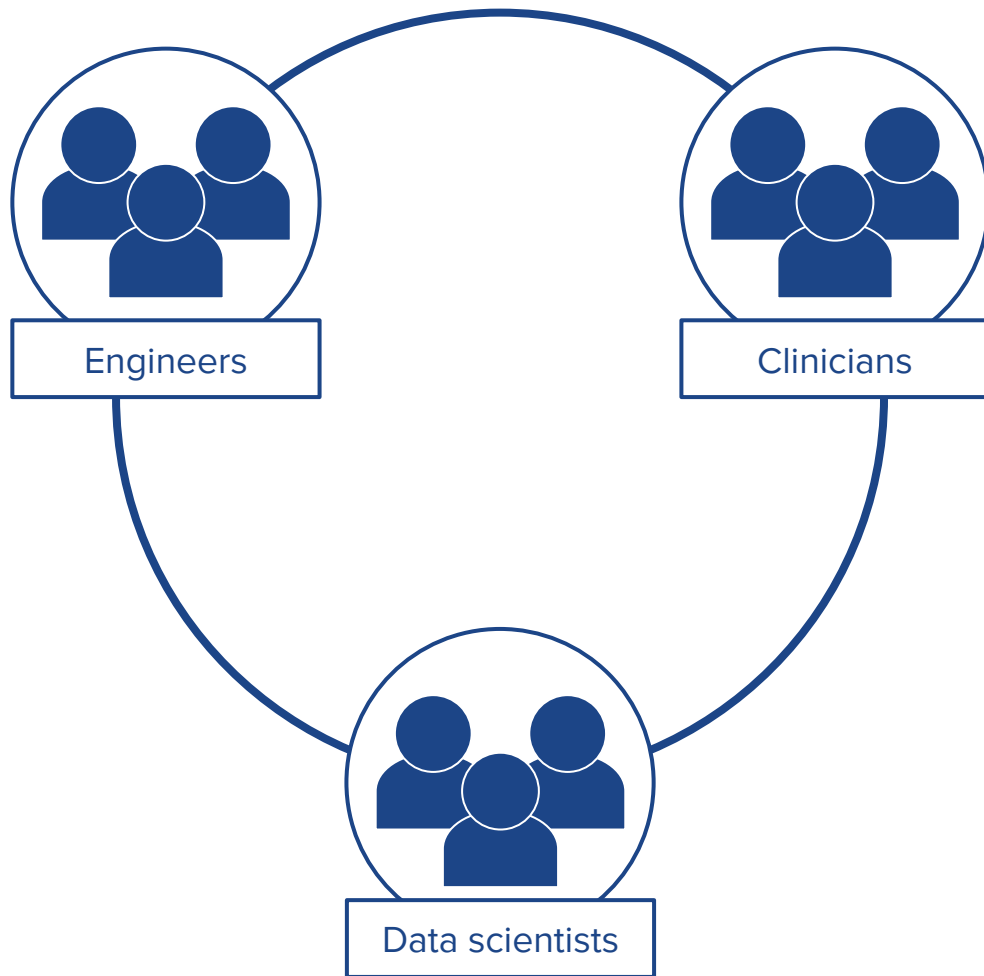
Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

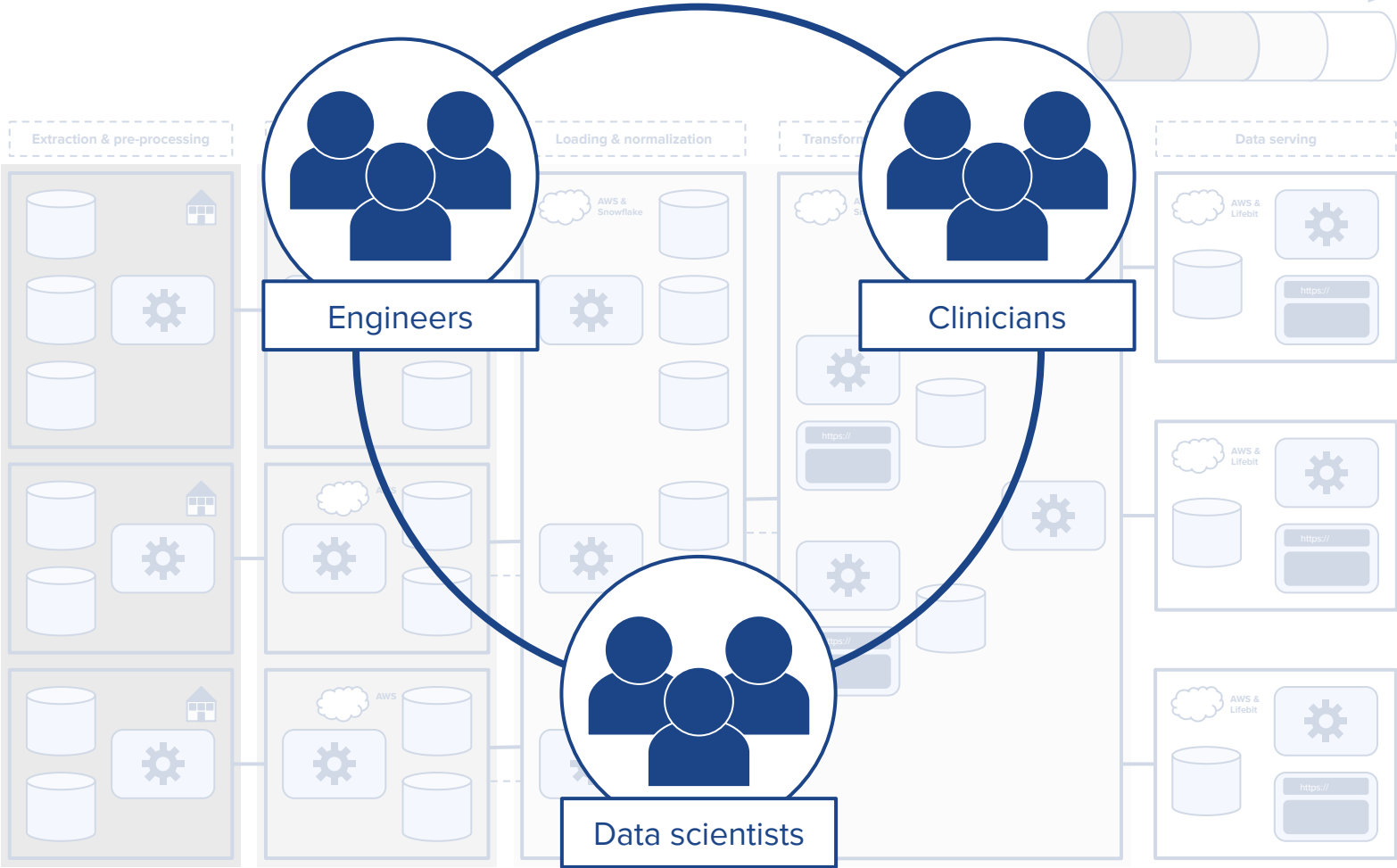
Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

What is real-world evidence (RWE) and why does it matter?

What makes it hard to generate RWE?

Feedback loops in our data architecture.

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

What is real-world evidence (RWE) and why does it matter?

What makes it hard to generate RWE?

Our technology stack as feedback machine.

Feedback loops in our data architecture.

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

-zsh

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

-zsh

```
poetry add dagster dagster-dbt
poetry add dbt-snowflake
poetry add "dlt[s3,snowflake]"
poetry add "ibis-framework[snowflake]"
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

-zsh

```
poetry add dagster dagster-dbt
poetry add dbt-snowflake
poetry add "dlt[s3,snowflake]"
poetry add "ibis-framework[snowflake]"
```

The impact on RWE
generation.

Patterns that accelerate
the loop.

Our technology stack
as feedback machine.

Feedback loops in our
data architecture.

What makes it hard to
generate RWE?

What is real-world
evidence (RWE) and
why does it matter?

-zsh

```
open "https://dagster.io"
```

Want to learn what it takes to build a data platform? Download the Data Platform Fundamentals e-book. [Learn more](#)



Product ▾

Solutions ▾

Pricing

Company ▾

Resources ▾

Try Dagster+

Sign In

Your platform for AI and data pipelines.

Dagster is a unified control plane for teams to build, scale, and observe their AI
& data pipelines with confidence.

Request a Demo

Try Dagster+

The impact on RWE
generation.

Patterns that accelerate
the loop.

Our technology stack
as feedback machine.

Feedback loops in our
data architecture.

What makes it hard to
generate RWE?

What is real-world
evidence (RWE) and
why does it matter?

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

-zsh

```
open "https://getdbt.com"
```


The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

getdbt.com



From CRN: dbt Labs among '15 Hottest AI Data and Analytics Companies'



Product ▾

Solutions ▾

Customers ▾

Resources ▾

Community ▾

Pricing



Try dbt

Book a demo



The dbt philosophy

The modern standard for data transformation.

dbt turns data work into a shared, scalable practice by bringing the best of software engineering to the analytics workflow.

Try dbt for free

Read customer stories



Raw Data

Analysis Ready

Enter

The impact on RWE
generation.

Patterns that accelerate
the loop.

Our technology stack
as feedback machine.

Feedback loops in our
data architecture.

What makes it hard to
generate RWE?

What is real-world
evidence (RWE) and
why does it matter?

-zsh

```
open "https://dlthub.com"
```

Lightweight Python code to move data

We focus on the needs & constraints of Python-first data platform teams: how to write any data source, achieve data democracy, modernise legacy systems and reduce cloud costs.

Get started

Get in touch

TRUSTED BY



PostHog



harness



Taktile

ARTSY

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

The impact on RWE
generation.

Patterns that accelerate
the loop.

Our technology stack
as feedback machine.

Feedback loops in our
data architecture.

What makes it hard to
generate RWE?

What is real-world
evidence (RWE) and
why does it matter?

-zsh

```
open "https://ibis-project.org"
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

ibis-project.org



Ibis

Getting started ▾

Concepts ▾

Backends ▾

How-to ▾

Reference ▾

Posts

Presentations

Release notes

Contribute ▾



Ibis

the portable Python dataframe library

Why Ibis?

Installation

10 minutes to Ibis

Blog

GitHub

Chat

RSS

An open source dataframe library that works with any data system

- Use the same API for nearly 20 backends
- Fast local dataframes with embedded DuckDB (default), Polars, or DataFusion
- Iterate locally and deploy remotely by changing a single line of code
- Compose SQL and Python dataframe code, bridging the gap between data engineering and data science

Ibis: the portable Python dataframe library

The impact on RWE
generation.

Patterns that accelerate
the loop.

Our technology stack
as feedback machine.

Feedback loops in our
data architecture.

What makes it hard to
generate RWE?

What is real-world
evidence (RWE) and
why does it matter?

-zsh

nano definitions.py

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

```
@dlt.transformer(standalone=True)
def bdt_transformer(items):
    ...

@dlt.source
def john_doe_clinic_bdt_dlt_source():
    return filesystem(bucket_url=...) | bdt_transformer()

@asset
def john_doe_clinic_bdt_dlt_source_asset(dlt: DltResource):
    source_schema = os.environ["JOHN_DOE_CLINIC_BDT_SOURCE_SCHEMA"]

    dlt.pipeline(source_schema).run(john_doe_clinic_bdt_dlt_source())

class DltResource(ConfigurableResource):
    def pipeline(self, source_schema):
        return dlt.pipeline(
            destination="snowflake",
            dataset_name=source_schema
        )
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

```
@dlt.transformer(standalone=True)
def bdt_transformer(items):
    ...

@dlt.source
def john_doe_clinic_bdt_dlt_source():
    return filesystem(bucket_url=...) | bdt_transformer()

@asset
def john_doe_clinic_bdt_dlt_source_asset(dlt: DltResource):
    source_schema = os.environ["JOHN_DOE_CLINIC_BDT_SOURCE_SCHEMA"]

    dlt.pipeline(source_schema).run(john_doe_clinic_bdt_dlt_source())

class DltResource(ConfigurableResource):
    def pipeline(self, source_schema):
        return dlt.pipeline(
            destination="snowflake",
            dataset_name=source_schema
        )
```


The impact on RWE
generation.

Patterns that accelerate
the loop.

Our technology stack
as feedback machine.

Feedback loops in our
data architecture.

What makes it hard to
generate RWE?

What is real-world
evidence (RWE) and
why does it matter?

-zsh

```
nano models/src/john_doe_clinic/bdt.yml
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

version: 2

sources:

- name: john_doe_clinic_bdt
database: '{{ env_var("SOURCE_DATABASE") }}'
schema: '{{ env_var("JOHN_DOE_CLINIC_BDT_SOURCE_SCHEMA") }}'
tables:
 - ...
 - name: _6200__6200
meta:
dagster:
asset_key: john_doe_clinic_bdt__6200__6200
 - name: _6200__6200__8401
meta:
dagster:
asset_key: john_doe_clinic_bdt__6200__6200__6200__8401
 - name: _6200__6200__8401__8410
meta:
dagster:
asset_key: john_doe_clinic_bdt__6200__6200__6200__8401__8410
 - ...

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

```
version: 2
```

```
sources:
```

```
- name: john_doe_clinic_bdt
  database: '{{ env_var("SOURCE_DATABASE") }}'
  schema: '{{ env_var("JOHN_DOE_CLINIC_BDT_SOURCE_SCHEMA") }}'
  tables:
    - ...
    - name: _6200__6200
      meta:
        dagster:
          asset_key: john_doe_clinic_bdt__6200__6200
    - name: _6200__6200__8401
      meta:
        dagster:
          asset_key: john_doe_clinic_bdt__6200__6200__6200__8401
    - name: _6200__6200__8401__8410
      meta:
        dagster:
          asset_key: john_doe_clinic_bdt__6200__6200__6200__8401__8410
    - ...
```

The impact on RWE
generation.

Patterns that accelerate
the loop.

Our technology stack
as feedback machine.

Feedback loops in our
data architecture.

What makes it hard to
generate RWE?

What is real-world
evidence (RWE) and
why does it matter?

-zsh

nano definitions.py

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

```
def asset_keys(dbt_source):
```

```
    ...
```

```
def john_doe_clinic_bdt_dbt_source_table_asset_factory():
```

```
    for asset_key in asset_keys("john_doe_clinic/bdt.yml"):
```

```
        @asset(name=asset_key, deps=[john_doe_clinic_bdt_dlt_source_asset])
```

```
        def dbt_source_table_asset():
```

```
            pass
```

```
        yield dbt_source_table_asset
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

```
def asset_keys(dbt_source):
```

```
    ...
```

```
def john_doe_clinic_bdt_dbt_source_table_asset_factory():
```

```
    for asset_key in asset_keys("john_doe_clinic/bdt.yml"):
```

```
        @asset(name=asset_key, deps=[john_doe_clinic_bdt_dlt_source_asset])
```

```
        def dbt_source_table_asset():
```

```
            pass
```

```
        yield dbt_source_table_asset
```

The impact on RWE
generation.

Patterns that accelerate
the loop.

Our technology stack
as feedback machine.

Feedback loops in our
data architecture.

What makes it hard to
generate RWE?

What is real-world
evidence (RWE) and
why does it matter?

-zsh

```
nano models/stg/john_doe_clinic/john_doe_clinic_lab.sql
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

```
select
```

```
    ...,
```

```
    to_date(cast(_8432__value as varchar), 'DDMMYYYY') as test_date,
```

```
    ...
```

```
from source("john_doe_clinic_bdt", "_6200__6200") as _6200
```

```
join source("john_doe_clinic_bdt", "_6200__6200__8401") as _8401
```

```
    on _6200._dlt_id = _8401._dlt_parent_id
```

```
left join source("john_doe_clinic_bdt", "_6200__6200__8401__8410") as _8410
```

```
    on _8401._dlt_id = _8410._dlt_parent_id
```


The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

```
select
    ...,
    to_date(cast(_8432__value as varchar), 'DDMMYYYY') as test_date,
    ...
from source("john_doe_clinic_bdt", "_6200__6200") as _6200
join source("john_doe_clinic_bdt", "_6200__6200__8401") as _8401
  on _6200._dlt_id = _8401._dlt_parent_id
left join source("john_doe_clinic_bdt", "_6200__6200__8401__8410") as _8410
  on _8401._dlt_id = _8410._dlt_parent_id
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

-zsh

```
poetry run dbt compile --profile="snowflake"
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

-zsh

```
poetry run dbt compile --profile="snowflake"
```

The impact on RWE
generation.

Patterns that accelerate
the loop.

Our technology stack
as feedback machine.

Feedback loops in our
data architecture.

What makes it hard to
generate RWE?

What is real-world
evidence (RWE) and
why does it matter?

-zsh

nano definitions.py

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

```
@dbt_assets(manifest=Path(...) / "manifest.json")
def dbt_model_assets(context: AssetExecutionContext, dbt: DbtCliResource):
    yield from dbt.cli(["build"], context=context).stream()
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

```
@dbt_assets(manifest=Path(...) / "manifest.json")
def dbt_model_assets(context: AssetExecutionContext, dbt: DbtCliResource):
    yield from dbt.cli(["build"], context=context).stream()
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

```
@asset(deps=[dbt_model_assets])
def john_doe_clinic_lab_harmonized_asset(ibis: IbisResource):
    with ibis.session() as session:
        table = session.table(...)

        for index, rows in enumerate(table.to_pandas_batches()):
            harmonized_rows = ...

            if index == 0:
                session.create_table(..., obj=harmonized_rows, overwrite=True)
            else:
                session.insert(..., obj=harmonized_rows, overwrite=False)

class IbisResource(ConfigurableResource):
    @contextmanager
    def session(self):
        connection = ibis.snowflake.connect(...)

        try:
            yield connection
        finally:
            connection.disconnect()
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

```
@asset(deps=[dbt_model_assets])
def john_doe_clinic_lab_harmonized_asset(ibis: IbisResource):
    with ibis.session() as session:
        table = session.table(...)

        for index, rows in enumerate(table.to_pandas_batches()):
            harmonized_rows = ...

            if index == 0:
                session.create_table(..., obj=harmonized_rows, overwrite=True)
            else:
                session.insert(..., obj=harmonized_rows, overwrite=False)

class IbisResource(ConfigurableResource):
    @contextmanager
    def session(self):
        connection = ibis.snowflake.connect(...)

        try:
            yield connection
        finally:
            connection.disconnect()
```


The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

```
defs = Definitions(  
    assets=[  
        john_doe_clinic_bdt_dlt_source_asset,  
        *john_doe_clinic_bdt_dbt_source_table_asset_factory(),  
        dbt_model_assets,  
        john_doe_clinic_lab_harmonized_asset,  
    ],  
    resources={  
        "dbt": DbtCliResource(project_dir=..., profile="snowflake"),  
        "dlt": DltResource(),  
        "ibis": IbisResource(),  
    },  
)
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

nano

```
defs = Definitions(  
    assets=[  
        john_doe_clinic_bdt_dlt_source_asset,  
        *john_doe_clinic_bdt_dbt_source_table_asset_factory(),  
        dbt_model_assets,  
        john_doe_clinic_lab_harmonized_asset,  
    ],  
    resources={  
        "dbt": DbtCliResource(project_dir=..., profile="snowflake"),  
        "dlt": DltResource(),  
        "ibis": IbisResource(),  
    },  
)
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

-zsh

```
poetry run dagster asset materialize -f definitions.py --select "*"
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

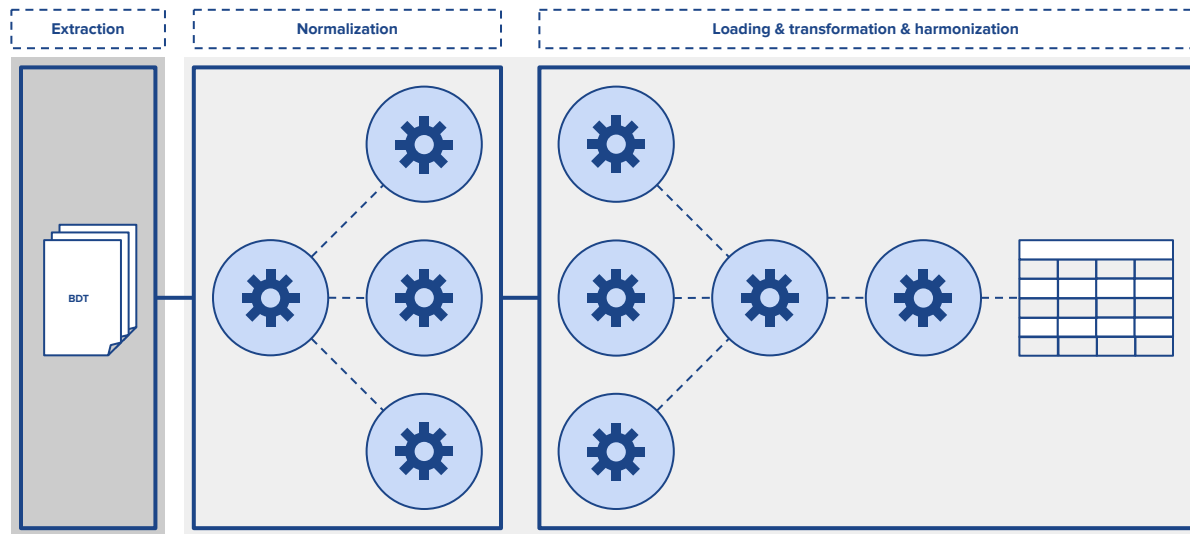
-zsh

```
poetry run dagster asset materialize -f definitions.py --select "*"

```

-zsh

```
poetry run dagster asset materialize -f definitions.py --select "*"
```



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

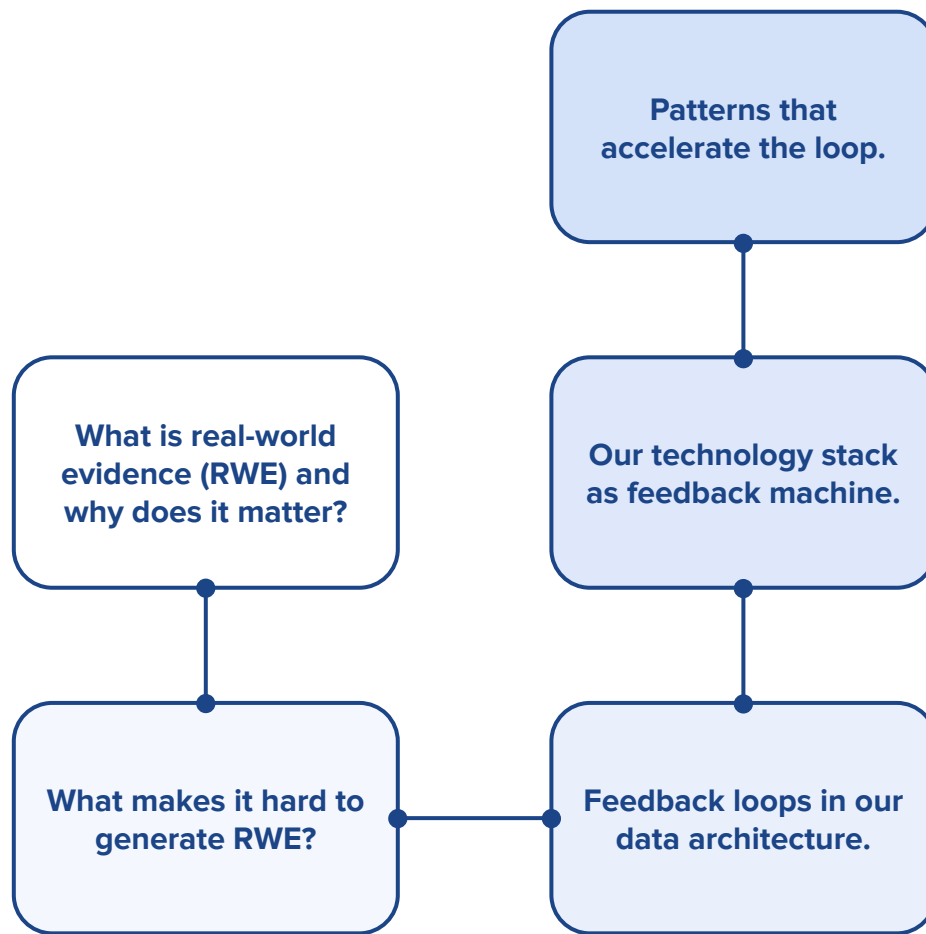
What is real-world evidence (RWE) and why does it matter?

What is real-world evidence (RWE) and why does it matter?

What makes it hard to generate RWE?

Our technology stack as feedback machine.

Feedback loops in our data architecture.



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers.

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for all scientists and engineers.

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

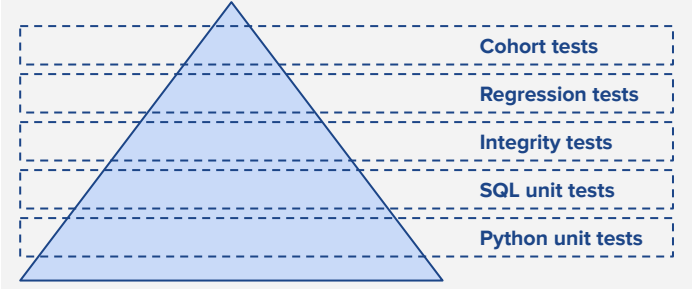
Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers.



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Drift tests	
Test prod	Drift test integrity
Tests, regression tests, unit tests	
Test features and endpoints	



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



- zsh

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



-zsh

```
nano models/stg/john_doe_clinic/john_doe_clinic_lab.yml
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



nano

version: 2

models:

- name: john_doe_clinic_lab

columns:

- name: test

data_tests:

- expect_row_count_quantiles_to_be_greater_than_or_equal_to:

filter: ...

cohort_filter: ...

row_count_quantiles: ...

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



nano

```
version: 2
```

```
models:
```

```
- name: john_doe_clinic_lab
```

```
  columns:
```

```
    - name: test
```

```
    data_tests:
```

```
      - expect_row_count_quantiles_to_be_greater_than_or_equal_to:
```

```
        filter: ...
```

```
        cohort_filter: ...
```

```
        row_count_quantiles: ...
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

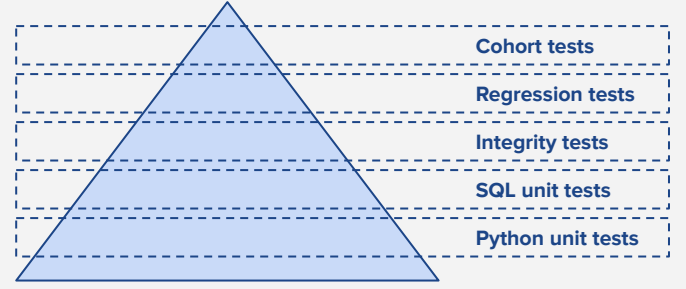
Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers.



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers.

Connect notebooks to pipelines.

- Prototype transformations that run in production.
- Fast feedback between clinicians and data scientists.

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers.

Connect notebooks to pipelines.

- Prototype transformations that run in production
 - Fast feedback between clinicians and data scientists.
- 

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.
<ul style="list-style-type: none">Test pyramid: Unit tests, integrity tests, regression tests, cohort testsFast feedback for data scientists and engineers

Connect to external services.
<ul style="list-style-type: none">Prototyping frameworks that run in productionFast feedback loops between dev and data scientists



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
Fast feedback for data scientists and engineers

Connect to data sources
Priority: performance that runs in production
Fast feedback for data scientists and data engineers



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
Fast feedback for data scientists and engineers

Connect to external services
Prototyping transforms that run in production
Fast feedback on schema and data access



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
<ul style="list-style-type: none">Test pyramid: Unit tests, integrity tests, regression tests, cohort testsFast feedback for data scientists and engineers

Connect to external systems
<ul style="list-style-type: none">Prototyping and performance that run in productionFast feedback on data science and data engineering



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

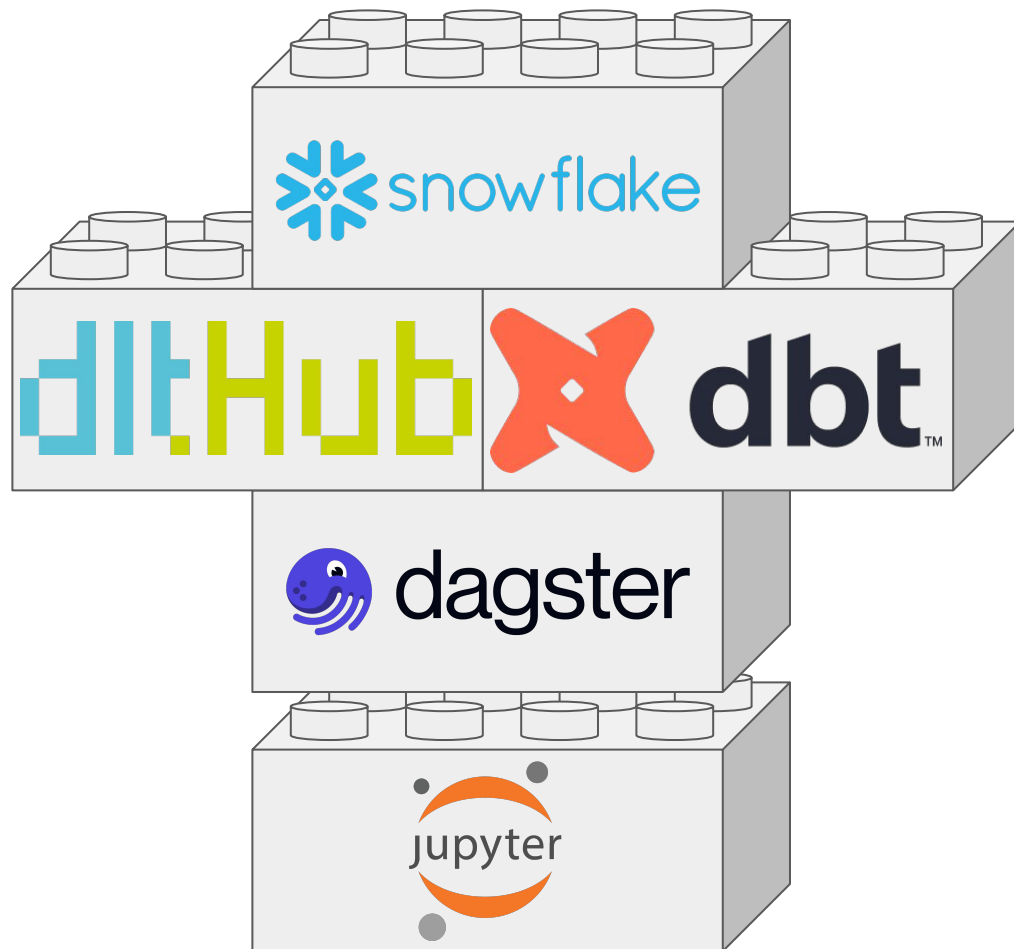
Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.
Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
Fast feedback for data scientists and engineers

Connect to external services.
Prototyping transforms that run in production
Fast feedback on transforms and data science



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

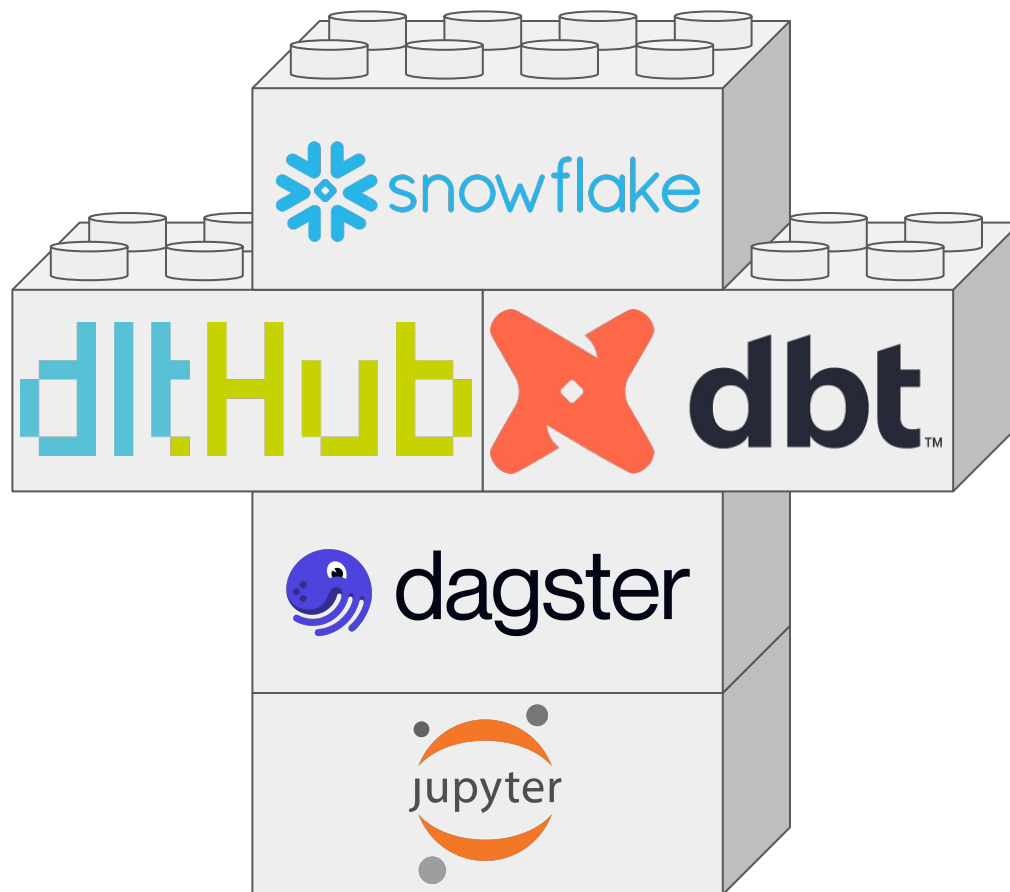
Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
Fast feedback for data scientists and engineers

Connect to data sources
Prototyping and deployment that run in production
Fast feedback for data scientists and data engineers



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.
<ul style="list-style-type: none">Test pyramid: Unit tests, integrity tests, regression tests, cohort testsFast feedback for data scientists and engineers

Connect to data sources.
<ul style="list-style-type: none">Prototyping transforms that run in productionFast feedback loops to data scientists and data engineers



- zsh

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
<ul style="list-style-type: none">Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
<ul style="list-style-type: none">Fast feedback for data scientists and engineers

Connect to external services
<ul style="list-style-type: none">Prototyping, transformation, that run in production
<ul style="list-style-type: none">Fast feedback for data scientists and data engineers



-zsh

```
poetry add ipykernel
poetry run python -m ipykernel install --user --name ... --display_name ...
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.
<ul style="list-style-type: none">Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
<ul style="list-style-type: none">Fast feedback for data scientists and engineers

Connect to external services.
<ul style="list-style-type: none">Prototyping, transformation, that run in production
<ul style="list-style-type: none">Fast feedback for data scientists and data engineers



-zsh

```
poetry add ipykernel
poetry run python -m ipykernel install --user --name ... --display_name ...
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
<ul style="list-style-type: none">Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
<ul style="list-style-type: none">Fast feedback for data scientists and engineers

Connect to data sources
<ul style="list-style-type: none">Prototyping, transformation, that run in production
<ul style="list-style-type: none">Fast feedback loops for data scientists and data engineers



-zsh

```
open "https://jupyter-4xzv.notebook.eu-central-1.sagemaker.aws/lab"
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
Fast feedback for data scientists and engineers

Connect to external services
Prototyping transformations that run in production
Fast feedback for data scientists and data engineers



jupyter-4xzv.notebook.eu-central-1.sagemaker.aws/lab

```
import dagster, definitions
```

```
@asset(deps=[definitions.dbt_model_assets])
def john_doe_clinic_lab_harmonized_asset(ibis: definitions.IbisResource):
    ...
```

```
dagster.materialize(
    assets=[
        definitions.john_doe_clinic_bdt_dlt_source_asset,
        *definitions.john_doe_clinic_bdt_dbt_source_table_asset_factory(),
        definitions.dbt_model_assets,
        john_doe_clinic_lab_harmonized_asset,
    ],
    resources={
        ...
    }
)
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
Fast feedback for data scientists and engineers

Connect to external services
Prototyping transforms that run in production
Fast feedback on transforms and data science



jupyter-4xzv.notebook.eu-central-1.sagemaker.aws/lab

```
import dagster, definitions
```

```
@asset(deps=[definitions.dbt_model_assets])
def john_doe_clinic_lab_harmonized_asset(ibis: definitions.IbisResource):
    ...
```

```
dagster.materialize(
    assets=[
        definitions.john_doe_clinic_bdt_dlt_source_asset,
        *definitions.john_doe_clinic_bdt_dbt_source_table_asset_factory(),
        definitions.dbt_model_assets,
        john_doe_clinic_lab_harmonized_asset,
    ],
    resources={
        ...
    }
)
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers.

Connect notebooks to pipelines.

- Prototype transformations that run in production
 - Fast feedback between clinicians and data scientists.
- 

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers.

Connect notebooks to pipelines.

- Prototype transformations that run in production.
- Fast feedback between clinicians and data scientists.

Make pipelines portable.

- Pipelines can fall back to local runtimes in restricted environments.
- Fast feedback between clinicians, data scientists and engineers.

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers.

Connect notebooks to pipelines.

- Prototype transformations that run in production.
- Fast feedback between clinicians and data scientists.

Make pipelines portable.

- Pipelines can fall back to local runtimes in restricted environments.
- Fast feedback between clinicians, data scientists and engineers.



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.
<ul style="list-style-type: none">Test pyramid: Unit tests, integrity tests, regression tests, cohort testsFast feedback for data scientists and engineers

Connect notebooks to pipelines.
<ul style="list-style-type: none">Prototype transformations that run in productionFast feedback between clinicians and data scientists

Make pipelines self-healing.
<ul style="list-style-type: none">Pipeline health: fit back, detect runtime errors, detect anomaliesFast feedback between clinicians, data scientists and engineers



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

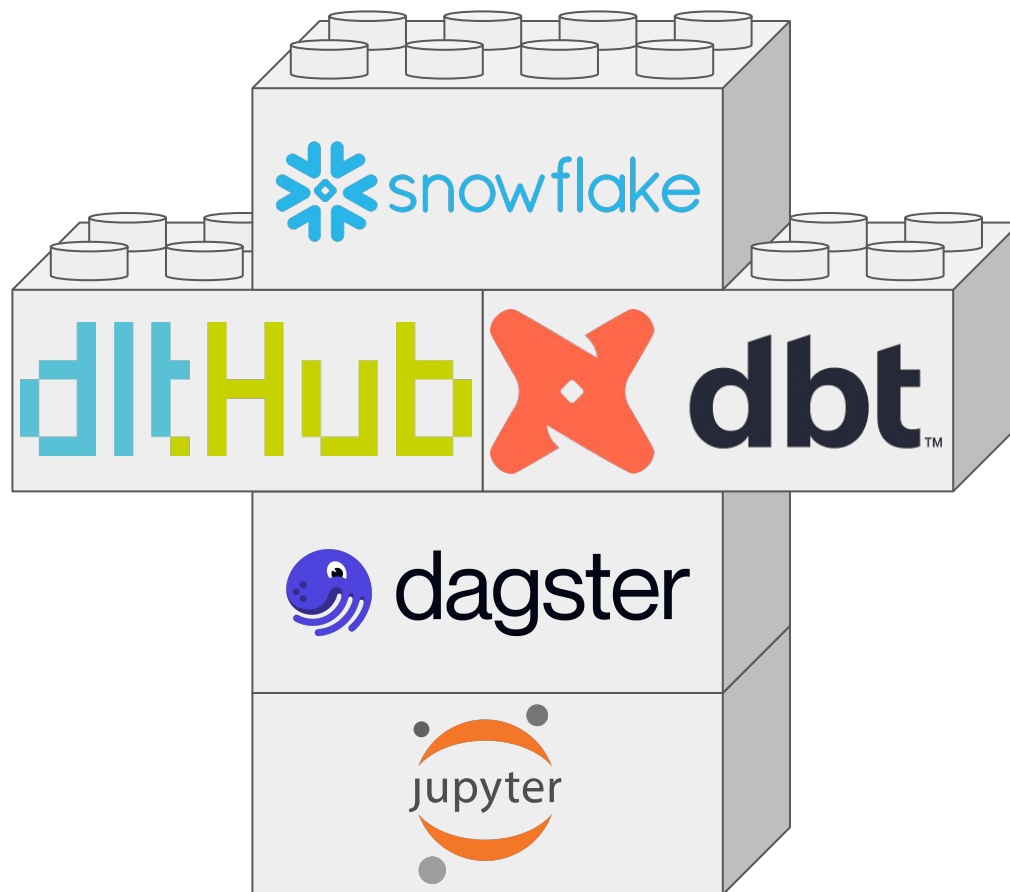
What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
<ul style="list-style-type: none">Test pyramids: unit tests, integrity tests, regression tests, cohort tests
<ul style="list-style-type: none">Fast feedback for data scientists and engineers

Connect notebooks to pipelines
<ul style="list-style-type: none">Prototype transformations that run in production
<ul style="list-style-type: none">Fast feedback between clinicians and data scientists

Make pipelines self-healing
<ul style="list-style-type: none">Provisioning for fast backfill and runtimes for critical workflows
<ul style="list-style-type: none">Fast feedback for data scientists and engineers



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
<ul style="list-style-type: none">Test pyramids: unit tests, integrity tests, regression tests, cohort tests
<ul style="list-style-type: none">Fast feedback for data scientists and engineers

Connect notebooks to pipelines
<ul style="list-style-type: none">Prototype transformations that run in production
<ul style="list-style-type: none">Fast feedback between clinicians and data scientists

Make pipelines self-healing
<ul style="list-style-type: none">Provisioning for fast feedback runtimes
<ul style="list-style-type: none">Fast feedback for data scientists and engineers



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

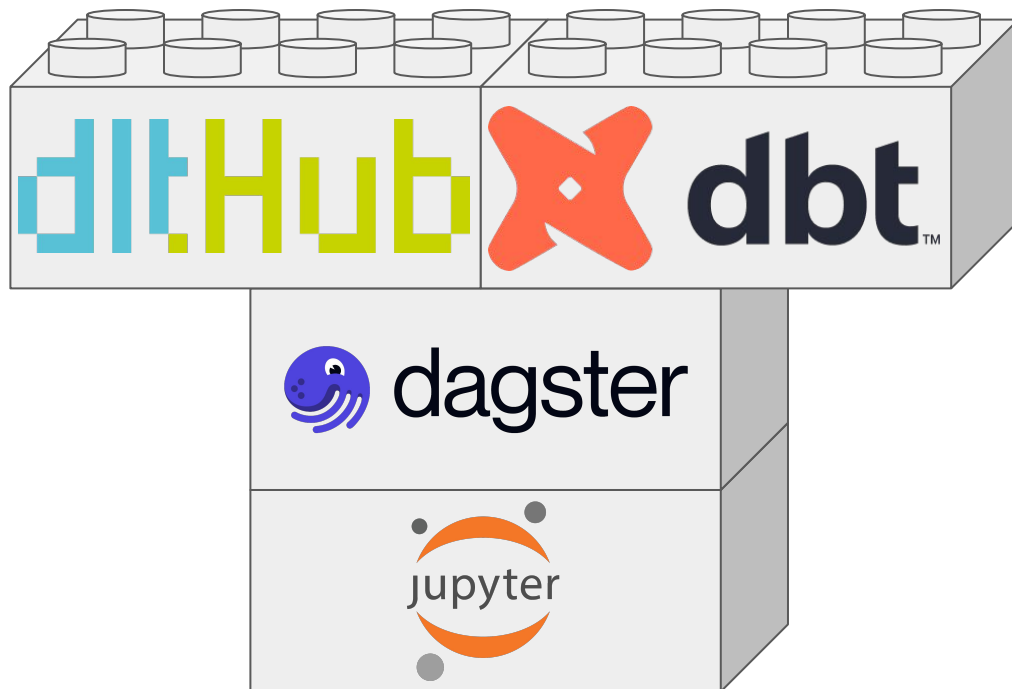
What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
<ul style="list-style-type: none">Test pyramid: unit tests, integrity tests, regression tests, cohort tests
<ul style="list-style-type: none">Fast feedback for data scientists and engineers

Connect notebooks to pipelines
<ul style="list-style-type: none">Prototype transformations that run in production
<ul style="list-style-type: none">Fast feedback between clinicians and data scientists

Make pipelines self-healing
<ul style="list-style-type: none">Provisioning for task-level runtimes / critical events
<ul style="list-style-type: none">Fast feedback for clinicians, data scientists and engineers



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

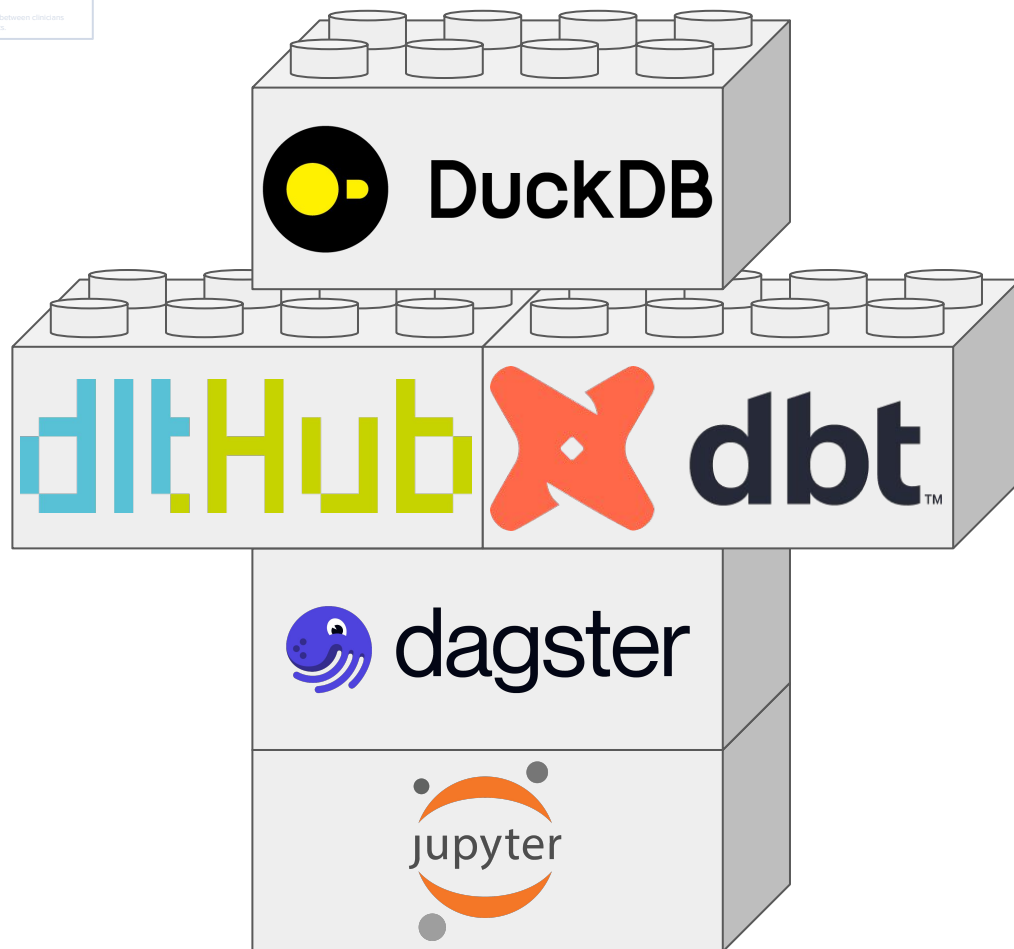
What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
<ul style="list-style-type: none">Test pyramids: unit tests, integrity tests, regression tests, cohort testsFast feedback for data scientists and engineers

Connect notebooks to pipelines
<ul style="list-style-type: none">Prototype transformations that run in productionFast feedback between clinicians and data scientists

Make pipelines self-healing
<ul style="list-style-type: none">Provisioning for fast backport and runtimesFast feedback for data scientists and engineers



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

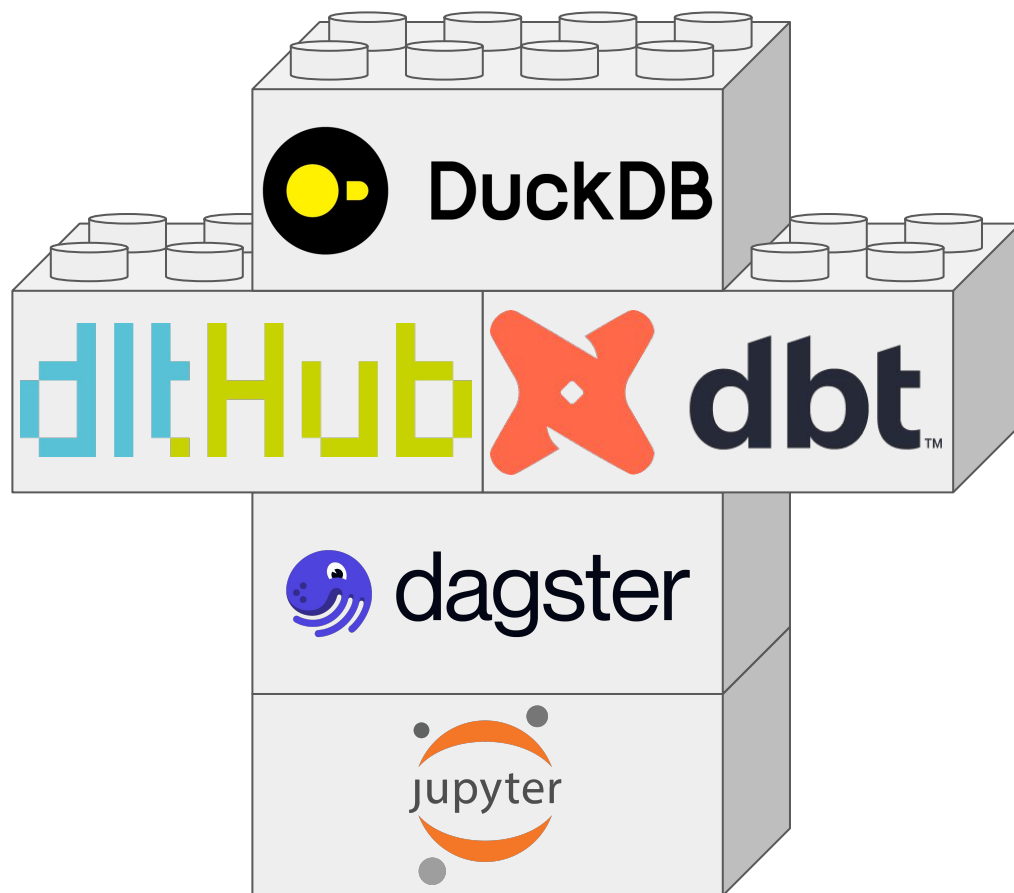
What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.
<ul style="list-style-type: none">Test pyramids: unit tests, integrity tests, regression tests, cohort testsFast feedback for data scientists and engineers

Connect notebooks to pipelines.
<ul style="list-style-type: none">Prototype transformations that run in productionFast feedback between clinicians and data scientists

Make pipelines self-healing.
<ul style="list-style-type: none">Provisioning for test, development, runtime, and production environmentsFast feedback for data scientists and engineers



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



- zsh

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers

Connect notebooks to pipelines.

- Prototype transformations that run in production
- Fast feedback between clinicians and data scientists

Make pipelines self-healing.

- Proactively detect backlogs, outages, runtime errors, blocked environments
- Fast feedback between data scientists and engineers

-zsh

```
poetry add dbt-duckdb
poetry add "dlt[duckdb,s3]"
poetry add "ibis-framework[duckdb]"
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



-zsh

```
poetry add dbt-duckdb
poetry add "dlt[duckdb,s3]"
poetry add "ibis-framework[duckdb]"
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers

Connect notebooks to pipelines.

- Prototype transformations that run in production
- Fast feedback between clinicians and data scientists

Make pipelines

- Pipelines in test back, local runtimes, trusted environments.
- Fast feedback between clinicians, data scientists and engineers



- zsh

open "https://duckdb.org"

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



duckdb.org



Documentation ▾

Resources ▾

GitHub ★ 29.6k

Support



We have released [DuckLake](#), a SQL-as-a-lakehouse standard. Read the [announcement blog post](#).



DuckDB is a fast analytical database system

Query and transform your data anywhere using DuckDB's feature-rich SQL dialect

Installation ▾

Documentation

SQL Python R Java Node.js

```
1 -- Get the top-3 busiest train stations
2 SELECT
3     station_name,
4     count(*) AS num_services
5 FROM train_services
6 GROUP BY ALL
7 ORDER BY num_services DESC
8 LIMIT 3;
```

Aggregation query ▾

Live demo →

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers

Connect notebooks to pipelines.

- Prototype transformations that run in production
- Fast feedback between clinicians and data scientists

Make pipelines self-healing.

- Pipeline health: full back, partial runtimes, critical error events.
- Fast feedback for data scientists and engineers



-zsh

nano definitions.py

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
<ul style="list-style-type: none">Test pyramids: Unit tests, integrity tests, regression tests, cohort tests
<ul style="list-style-type: none">Fast feedback for data scientists and engineers

Connect notebooks to pipelines
<ul style="list-style-type: none">Prototype transformations that run in production
<ul style="list-style-type: none">Fast feedback between clinicians and data scientists

Make pipelines run faster
<ul style="list-style-type: none">Parallelize pipelines: split tasks, parallel runtimes, distributed environments
<ul style="list-style-type: none">Fast feedback for data scientists and engineers



nano

```
class DltResource(ConfigurableResource):
    def pipeline(self, source_schema):
        return dlt.pipeline(
            destination="duckdb",
            dataset_name=source_schema
        )
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers

Connect notebooks to pipelines

- Prototype transformations that run in production
- Fast feedback between clinicians and data scientists

Make pipelines reusable

- Pipelines in full back-end runtimes (cloud or on-prem), tracked and governed
- Fast feedback between clinicians, data scientists and engineers



nano

```
class DltResource(ConfigurableResource):  
    def pipeline(self, source_schema):  
        return dlt.pipeline(  
            destination="duckdb",  
            dataset_name=source_schema  
        )
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.
<ul style="list-style-type: none">Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
<ul style="list-style-type: none">Fast feedback for data scientists and engineers

Connect notebooks to pipelines.
<ul style="list-style-type: none">Prototype transformations that run in production
<ul style="list-style-type: none">Fast feedback between clinicians and data scientists

Make pipelines reusable.
<ul style="list-style-type: none">Pipelines as first-class architectural elements.
<ul style="list-style-type: none">Fast feedback between clinicians, data scientists and engineers



-zsh

```
nano models/stg/john_doe_clinic/john_doe_clinic_lab.sql
```


The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
<ul style="list-style-type: none">Test pyramids: unit tests, integrity tests, regression tests, cohort testsFast feedback for data scientists and engineers

Connect notebooks to pipelines
<ul style="list-style-type: none">Prototype transformations that run in productionFast feedback between clinicians and data scientists

Make pipelines reusable
<ul style="list-style-type: none">Provisioning for fast, repeat runtimesEnforced security and access controlsFast feedback between clinicians, data scientists and engineers



nano

```
select
```

```
    ...,
```

```
    {{ to_date("_8432__value", "DDMMYYYY") }} as test_date,
```

```
    ...
```

```
from source("john_doe_clinic_bdt", "_6200__6200") as _6200
```

```
join source("john_doe_clinic_bdt", "_6200__6200__8401") as _8401
```

```
    on _6200._dlt_id = _8401._dlt_parent_id
```

```
left join source("john_doe_clinic_bdt", "_6200__6200__8401__8410") as _8410
```

```
    on _8401._dlt_id = _8410._dlt_parent_id
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers

Connect notebooks to pipelines

- Prototype transformations that run in production
- Fast feedback between clinicians and data scientists

Make pipelines self-healing

- Pipeline health: full back, partial runtimes, blocked on dependencies
- Fast feedback for data scientists and engineers

nano

```
select
```

```
    .../
```

```
    {{ to_date("_8432__value", "DDMMYYYY") }} as test_date,
```

```
    ...
```

```
from source("john_doe_clinic_bdt", "_6200__6200") as _6200
```

```
join source("john_doe_clinic_bdt", "_6200__6200__8401") as _8401
```

```
    on _6200._dlt_id = _8401._dlt_parent_id
```

```
left join source("john_doe_clinic_bdt", "_6200__6200__8401__8410") as _8410
```

```
    on _8401._dlt_id = _8410._dlt_parent_id
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers

Connect notebooks to pipelines.

- Prototype transformations that run in production
- Fast feedback between clinicians and data scientists

Make pipelines

- Pipelines in test back, control runtimes, limited environments
- Fast feedback between clinicians, data scientists and engineers



-zsh

nano macros/dialects/to_date.sql

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



nano

```
{% macro to_date(column, format) %}
    {{ return(adapter.dispatch("to_date")(column, format)) }}
{% endmacro %}

{% macro duckdb__to_date(column, format) %}
    {% set strpformat = format %}
    {% set strpformat = strpformat | replace("YYYY", "%Y") %}
    {% set strpformat = strpformat | replace("MM", "%m") %}
    {% set strpformat = strpformat | replace("DD", "%d") %}

    cast(strptime(cast('{{ column }} as varchar), '{{ strpformat }}') as date)
{% endmacro %}
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
<ul style="list-style-type: none">Test pyramid: Unit tests, integrity tests, regression tests, cohort testsFast feedback for data scientists and engineers

Connect notebooks to pipelines
<ul style="list-style-type: none">Prototype transformations that run in productionFast feedback between clinicians and data scientists

Make pipelines self-healing
<ul style="list-style-type: none">Proactively detect backlogs, outages, runtime errors, blocked dependenciesFast feedback to data scientists and engineers



nano

```
{% macro to_date(column, format) %}
    {{ return(adapter.dispatch("to_date")(column, format)) }}
{% endmacro %}

{% macro duckdb__to_date(column, format) %}
    {% set strpformat = format %}
    {% set strpformat = strpformat | replace("YYYY", "%Y") %}
    {% set strpformat = strpformat | replace("MM", "%m") %}
    {% set strpformat = strpformat | replace("DD", "%d") %}

    cast(strptime(cast('{{ column }}' as varchar), '{{ strpformat }}') as date)
{% endmacro %}
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers

Connect notebooks to pipelines.

- Prototype transformations that run in production
- Fast feedback between clinicians and data scientists

Make pipelines

- Pipelines in the back-end: real-time, batch, and streaming
- Fast feedback for data scientists and engineers



-zsh

```
poetry run dbt compile --profile="duckdb"
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers

Connect notebooks to pipelines.

- Prototype transformations that run in production
- Fast feedback between clinicians and data scientists

Make pipelines

- Provisioned in test backends, runtimes, and deployed environments.
- Fast feedback between clinicians, data scientists and engineers



-zsh

```
poetry run dbt compile --profile="duckdb"
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers

Connect notebooks to pipelines.

- Prototype transformations that run in production
- Fast feedback between clinicians and data scientists

Make pipelines

- Pipelines in the back-end, not runtimes in the front-end
- Fast feedback between clinicians, data scientists and engineers



-zsh

nano definitions.py

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



-zsh

```
class IbisResource(ConfigurableResource):
    @contextmanager
    def session(self):
        connection = ibis.duckdb.connect(...)

        try:
            yield connection
        finally:
            connection.disconnect()
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left.

- Test pyramid: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers

Connect notebooks to pipelines.

- Prototype transformations that run in production
- Fast feedback between clinicians and data scientists

Make pipelines run in the background.

- Pushdown to the back-end: runtimes, distributed engines
- Fast feedback between clinicians, data scientists and engineers



-zsh

```
class IbisResource(ConfigurableResource):
    @contextmanager
    def session(self):
        connection = ibis.duckdb.connect(...)

        try:
            yield connection
        finally:
            connection.disconnect()
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left
<ul style="list-style-type: none">Test pyramids: unit tests, integrity tests, regression tests, cohort testsFast feedback for data scientists and engineers

Connect notebooks to pipelines
<ul style="list-style-type: none">Prototype transformations that run in productionFast feedback between clinicians and data scientists

Make pipelines reusable
<ul style="list-style-type: none">Provisioning: test, dev, prod, runtimeOrchestration: scheduled or triggered workflowsFast feedback between clinicians, data scientists and engineers



-zsh

```
defs = Definitions(  
    assets=[  
        john_doe_clinic_bdt_dlt_source_asset,  
        *john_doe_clinic_bdt_dbt_source_table_asset_factory(),  
        dbt_model_assets,  
        john_doe_clinic_lab_harmonized_asset,  
    ],  
    resources={  
        "dbt": DbtCliResource(project_dir=..., profile="duckdb"),  
        "dlt": DltResource(),  
        "ibis": IbisResource(),  
    },  
    executor=in_process_executor  
)
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



-zsh

```
defs = Definitions(  
    assets=[  
        john_doe_clinic_bdt_dlt_source_asset,  
        *john_doe_clinic_bdt_dbt_source_table_asset_factory(),  
        dbt_model_assets,  
        john_doe_clinic_lab_harmonized_asset,  
    ],  
    resources={  
        "dbt": DbtCliResource(project_dir=..., profile="duckdb"),  
        "dlt": DltResource(),  
        "ibis": IbisResource(),  
    },  
    executor=in_process_executor  
)
```

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Shift testing left:

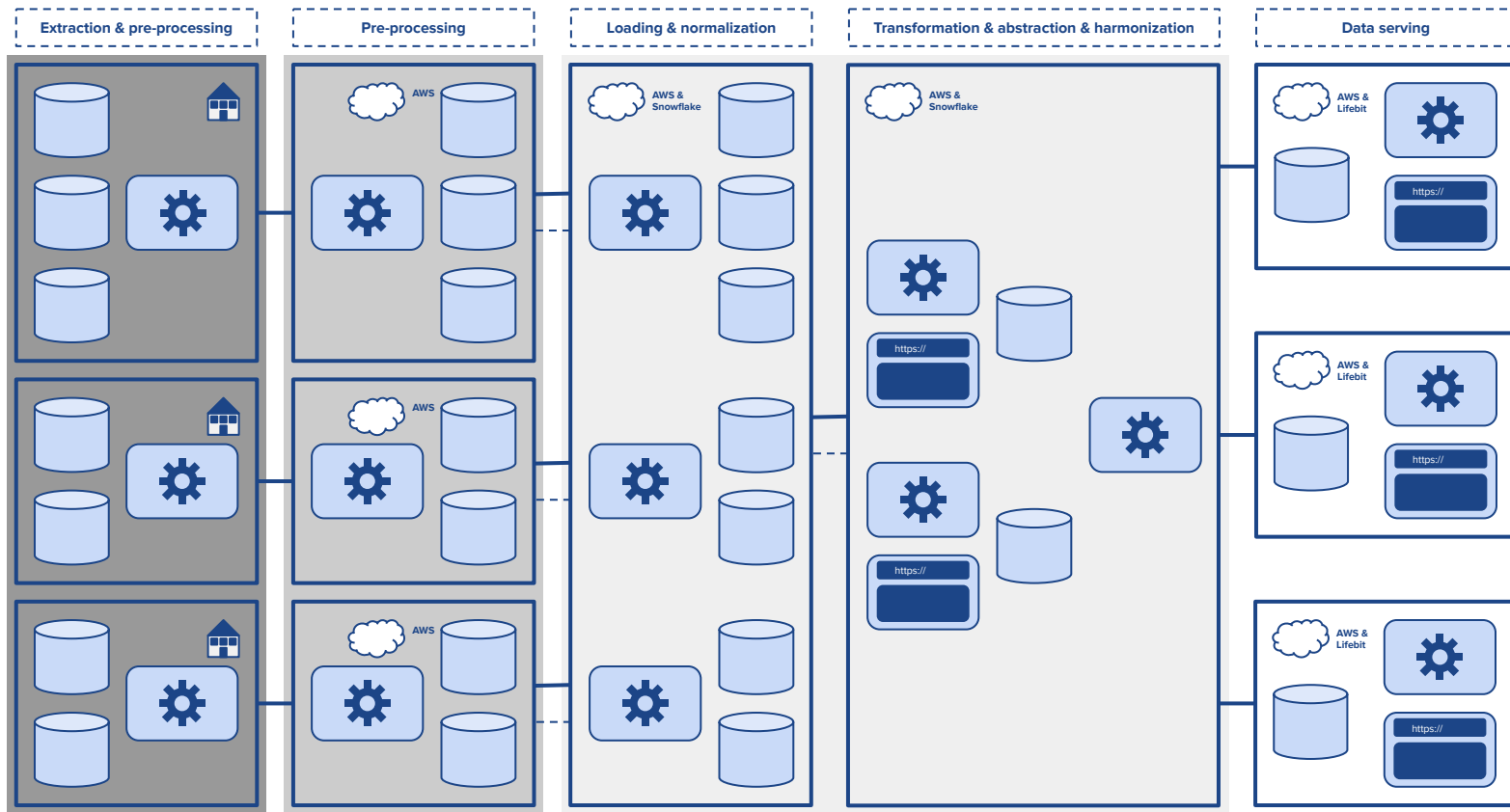
- Test pyramids: Unit tests, integrity tests, regression tests, cohort tests
- Fast feedback for data scientists and engineers

Connect notebooks to pipelines:

- Prototype transformations that run in production
- Fast feedback between clinicians and data scientists

Make pipelines
Provenance for data lineage and
auditability

Increasing privacy levels



The impact on RWE generation.

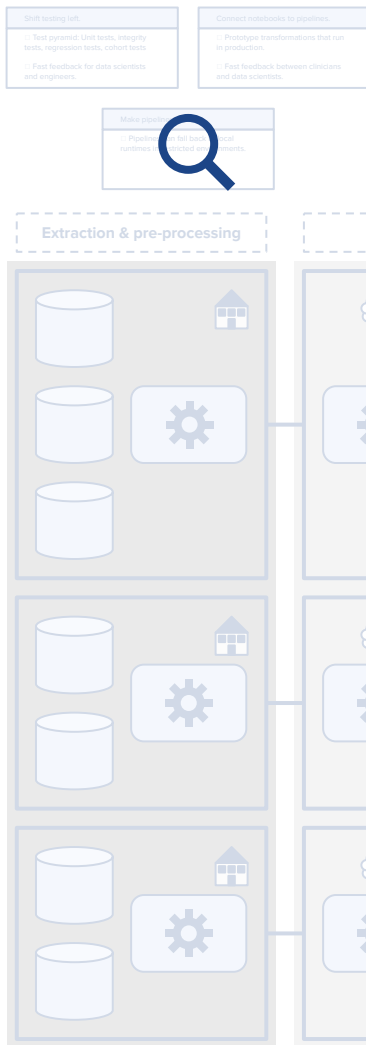
Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

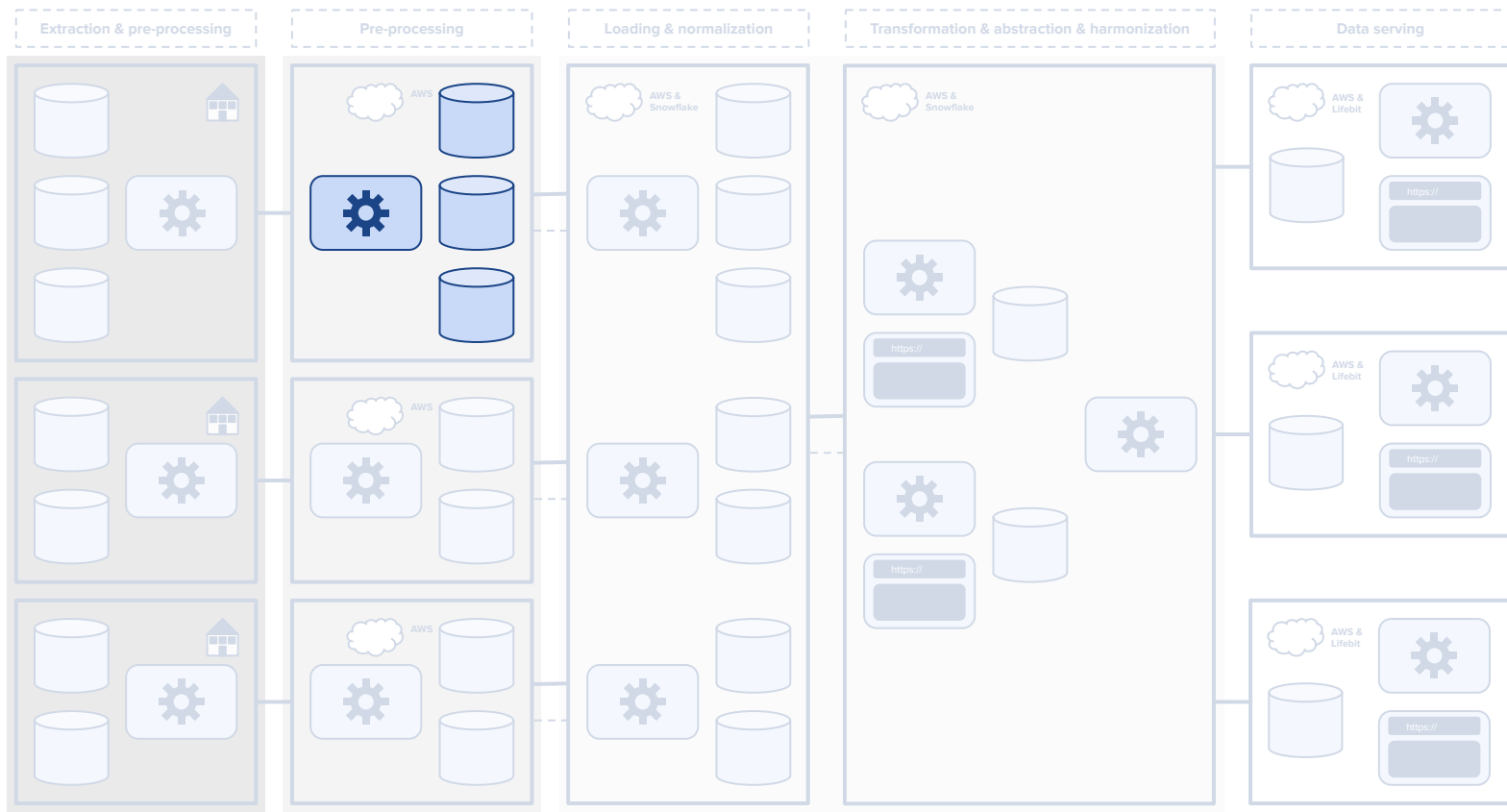
Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?



The impact on RWE generation.

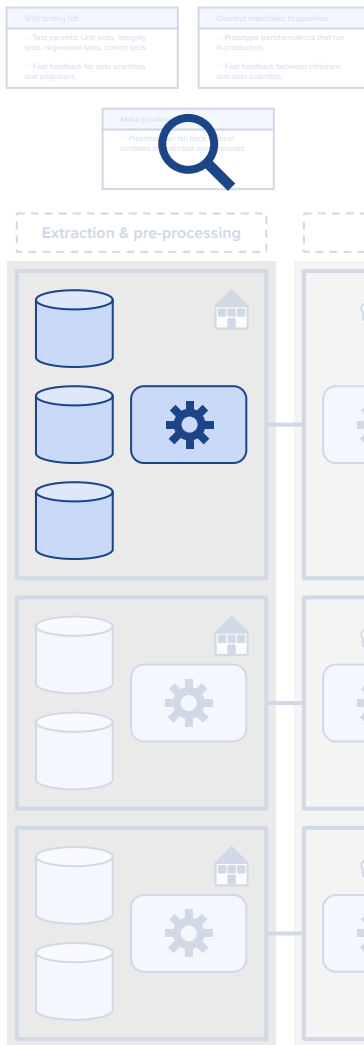
Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

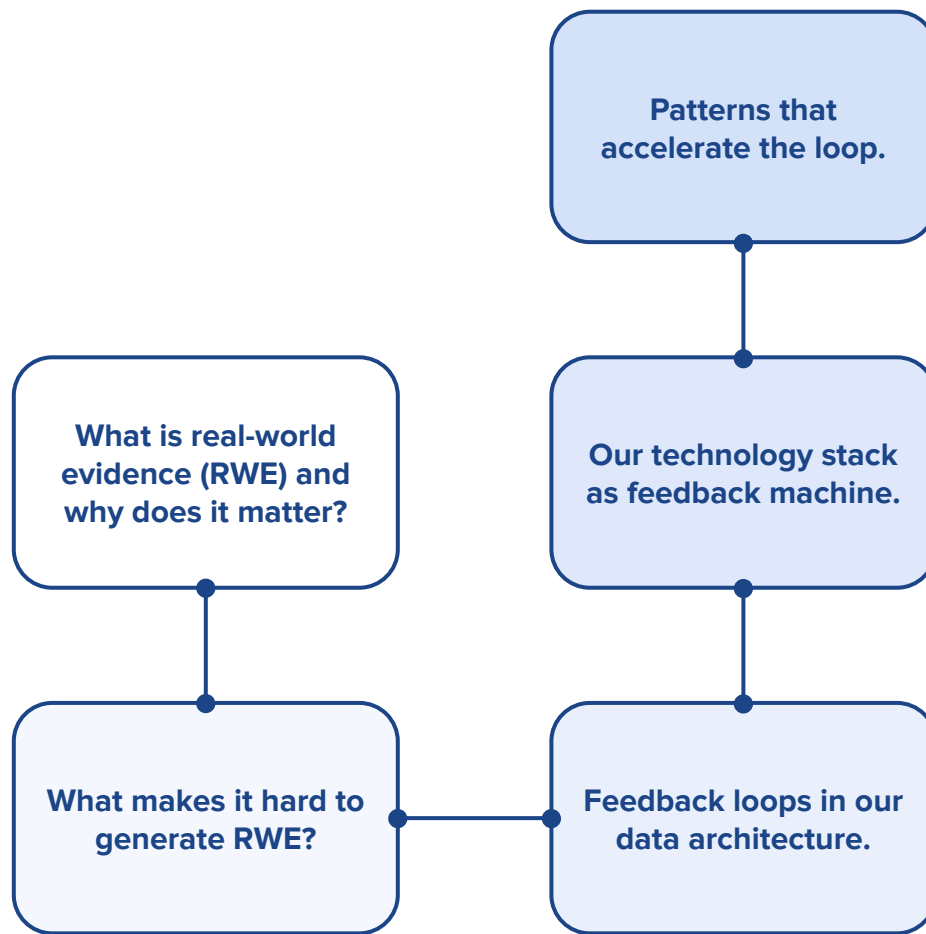
What makes it hard to generate RWE?

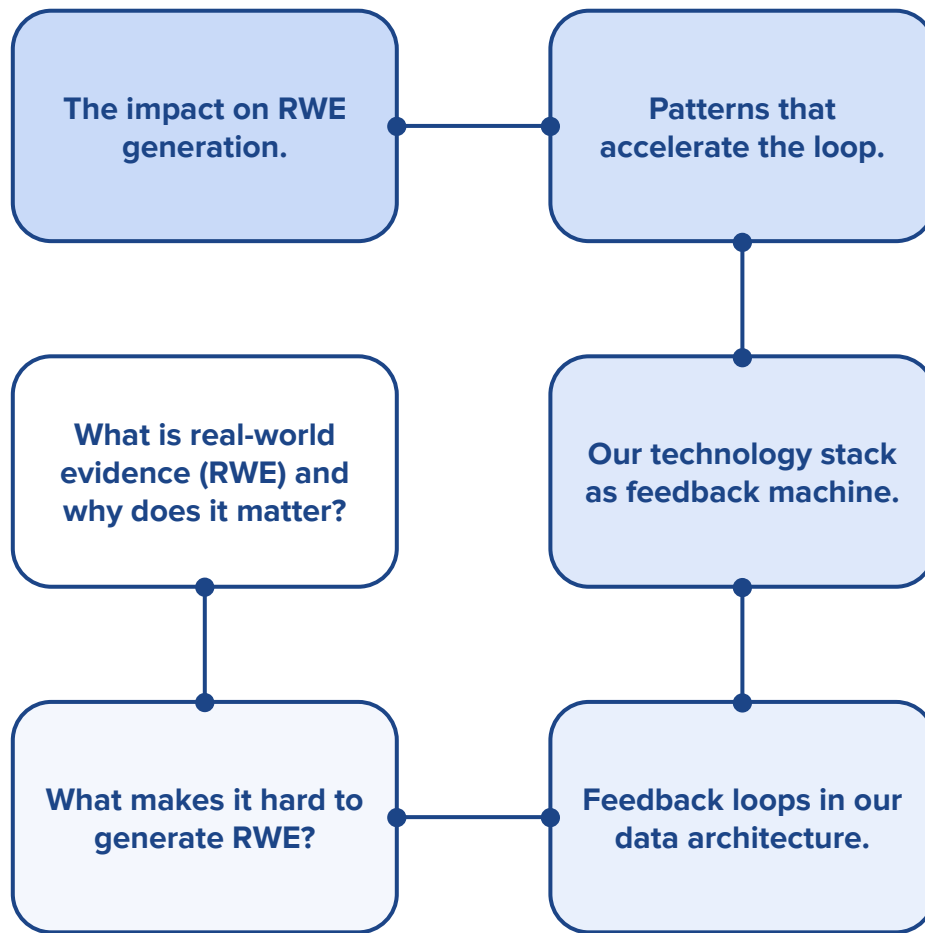
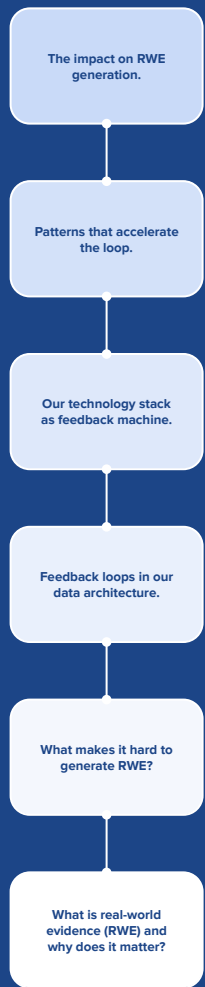
What is real-world evidence (RWE) and why does it matter?



Increasing privacy levels







The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

RWE at scale

- Our data has been used in 1,200+ publications
- Research using our data has been cited 9000+ times across peer-reviewed publications
- Flatiron has published research on 25+ cancer types

The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

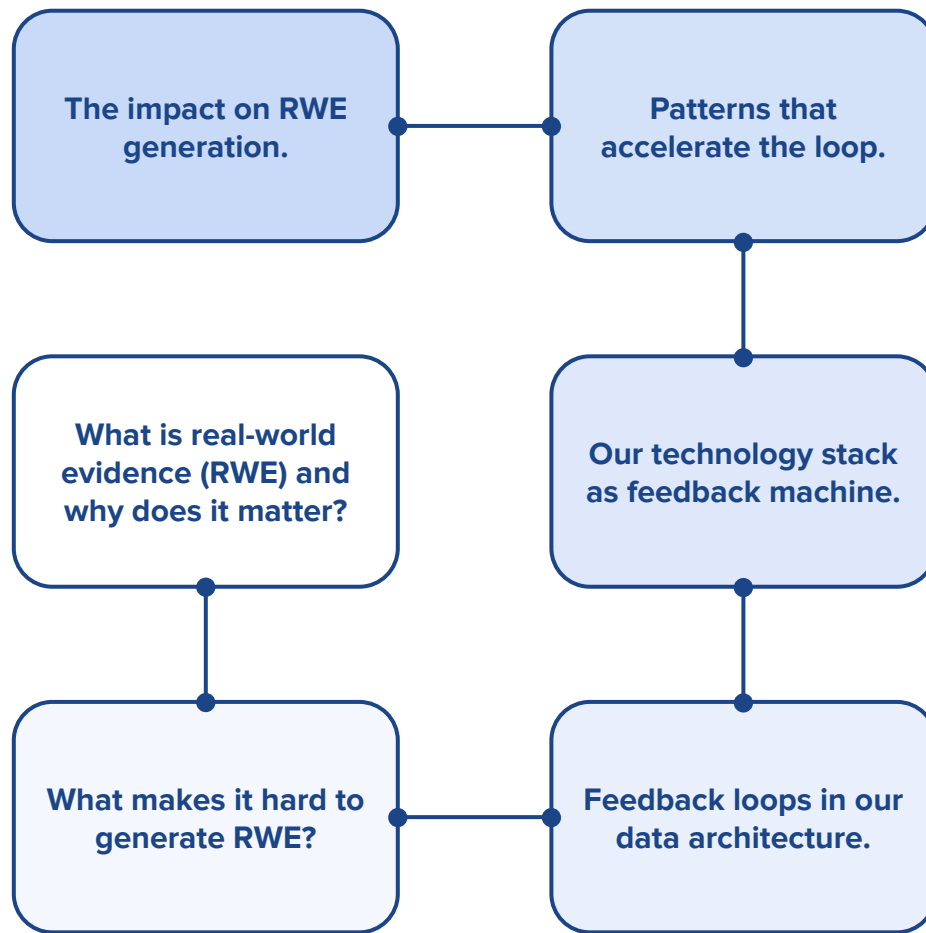
What is real-world evidence (RWE) and why does it matter?

RWE at scale

- Our data has been used in 1,200+ publications
- Research using our data has been cited 9000+ times across peer-reviewed publications
- Flatiron has published research on 25+ cancer types

Global reach

“The methods (..) enable global researchers with the tools to investigate treatment patterns and outcomes across diverse healthcare systems, fostering high-quality, multinational evidence generation - all while working within local regulatory requirements. This work (...) [helps to] accelerate the development of therapies, improve treatment strategies, and most importantly, **benefit patients worldwide by ensuring that high-quality RWE informs clinical decisions and health policies on a global scale.**”



The impact on RWE generation.

Patterns that accelerate the loop.

Our technology stack as feedback machine.

Feedback loops in our data architecture.

What makes it hard to generate RWE?

What is real-world evidence (RWE) and why does it matter?

Thank you!