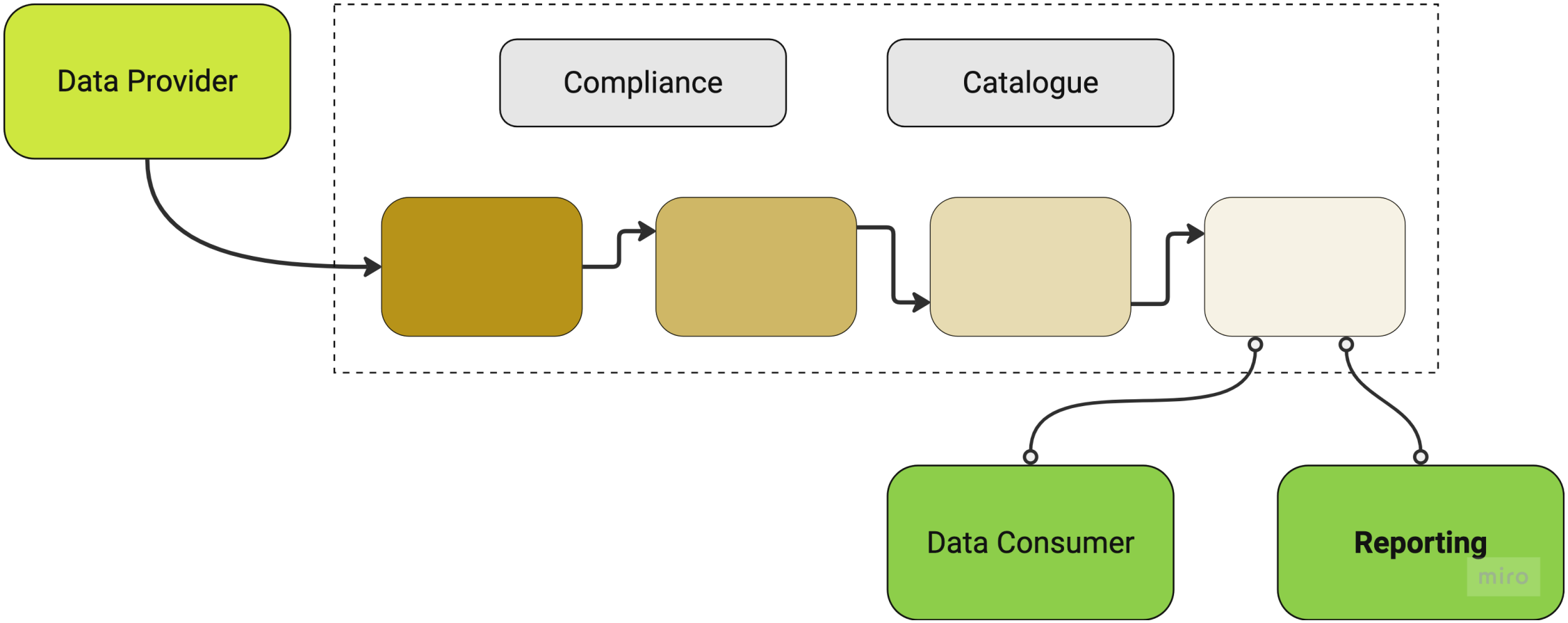
Applying engineering principles to reporting

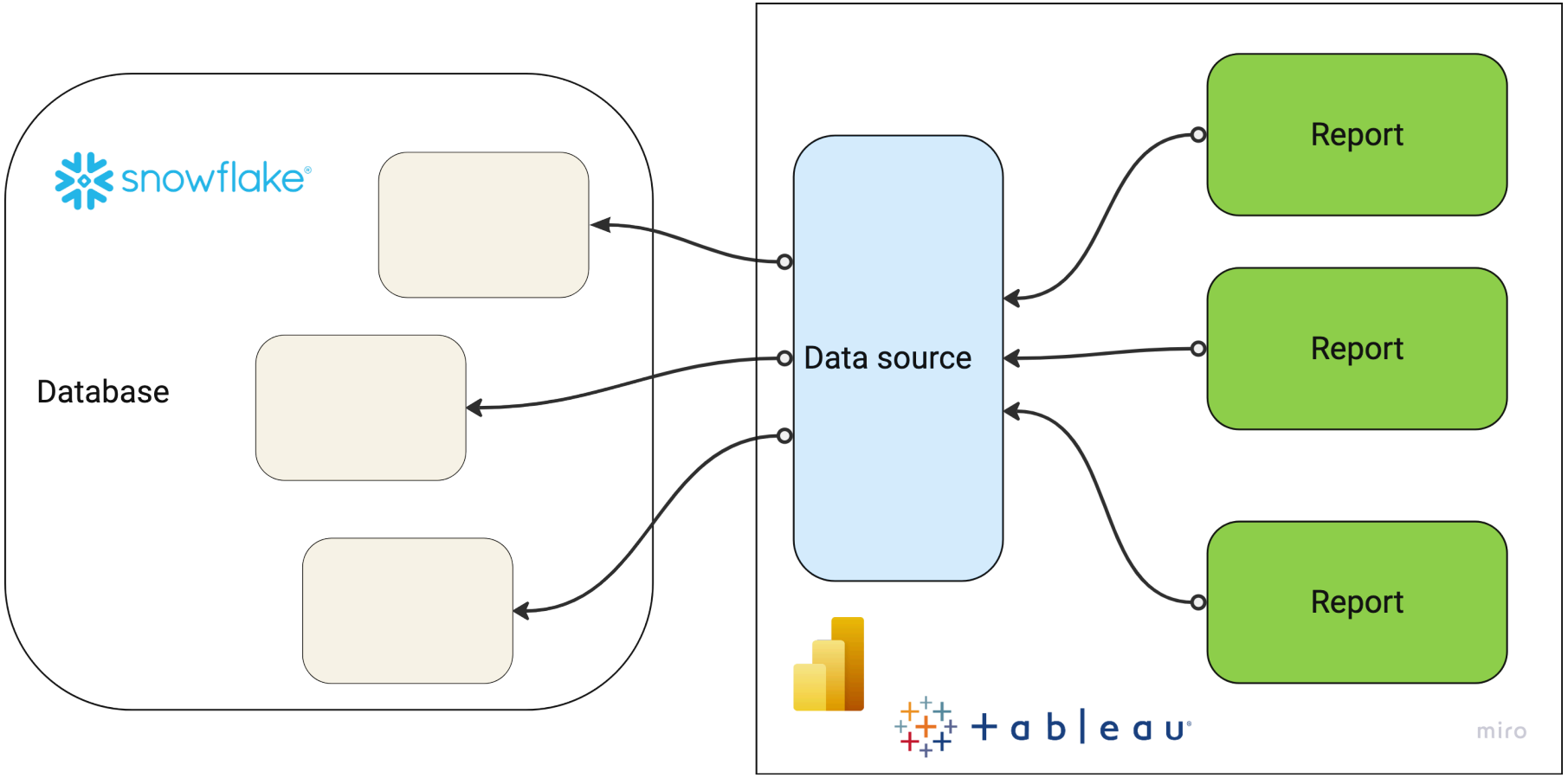
Guillaume (**G**) Belrose

Data Pipelines: What | How



EQUAL EXPERTS

Reporting 101



EQUAL EXPERTS

Agenda

Discovery, extend right & shift left

Discovery

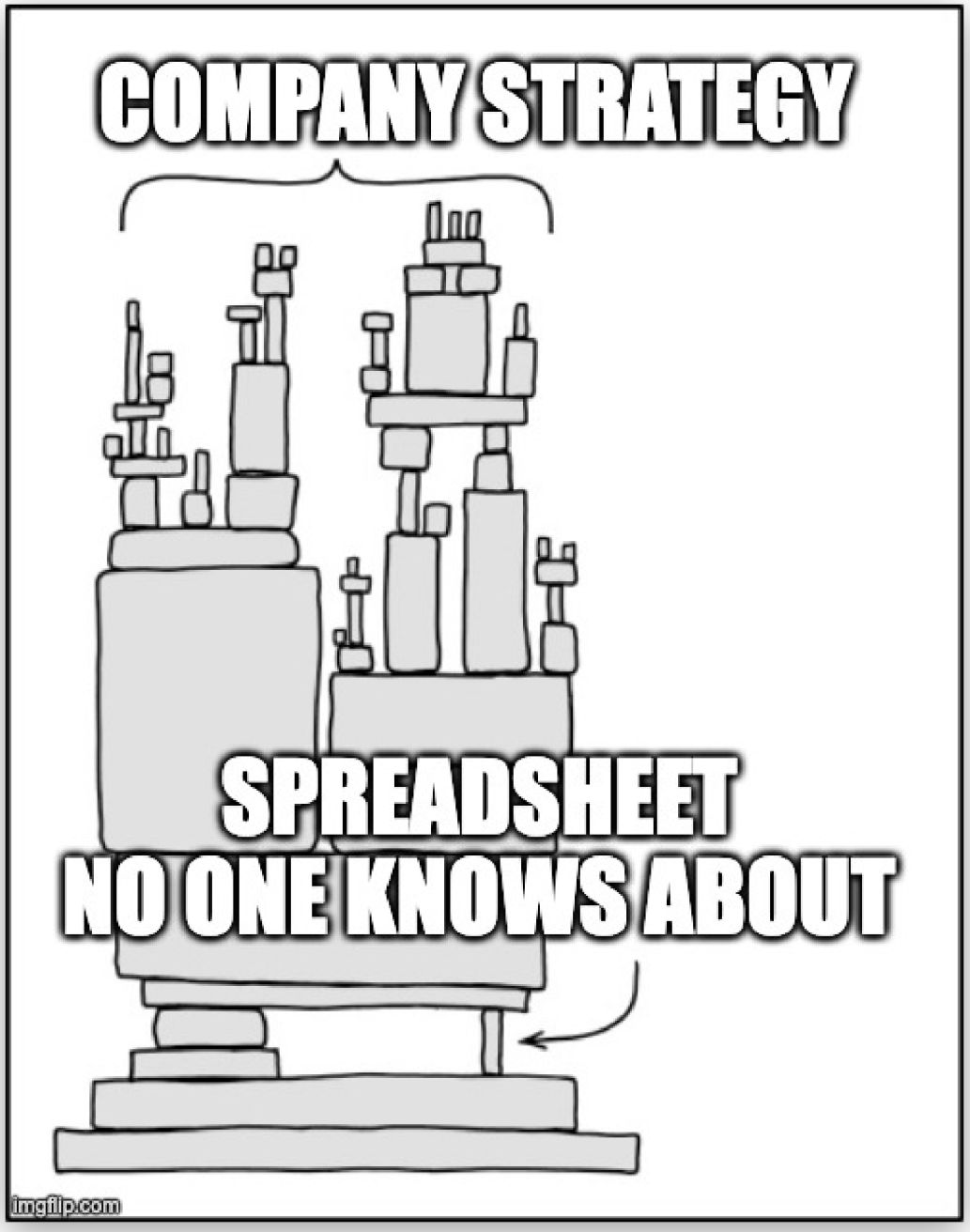
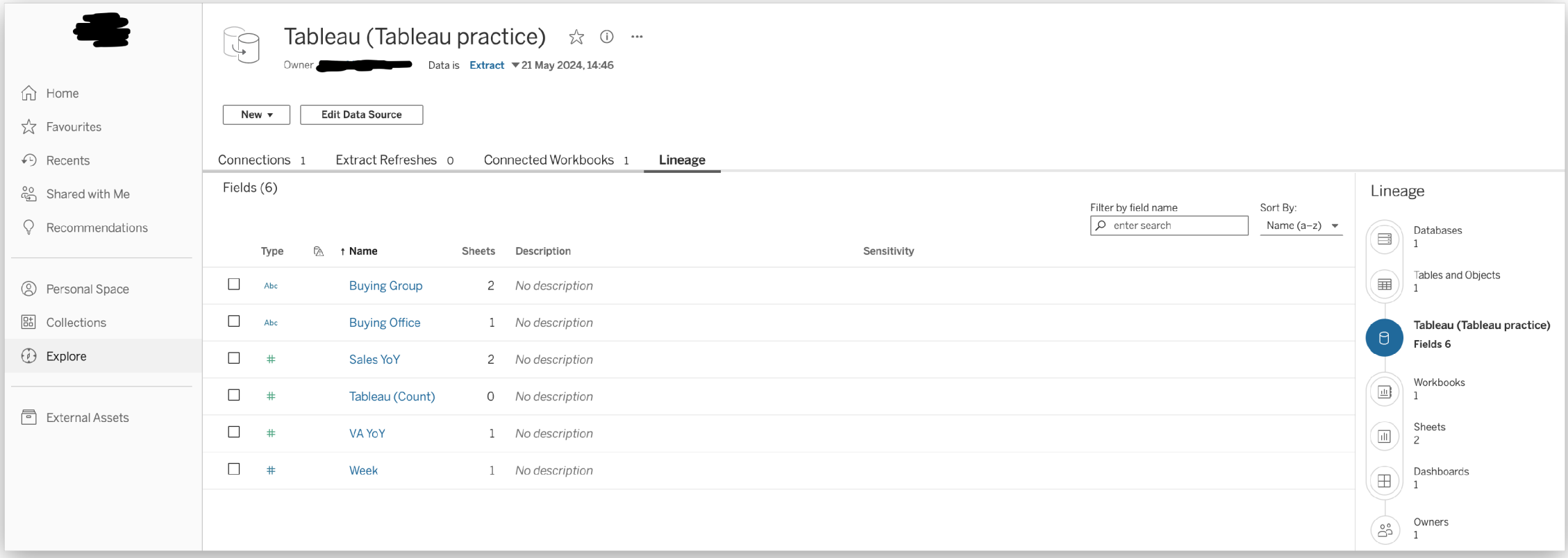

COMPANY STRATEGY
SPREADSHEET
NO ONE KNOWS ABOUT
imgflip.com

Tableau lineage



EQUAL EXPERTS

Tableau lineage on steroids

```
query DatasourceDetails($projectName: String!){
  publishedDatasources(filter:{projectName: $projectName}){
    luid
    name
    downstreamWorkbooks{
      luid
      name
    }
  }
}
```

See: Tableau Metadata API, GraphQL

EQUAL EXPERTS

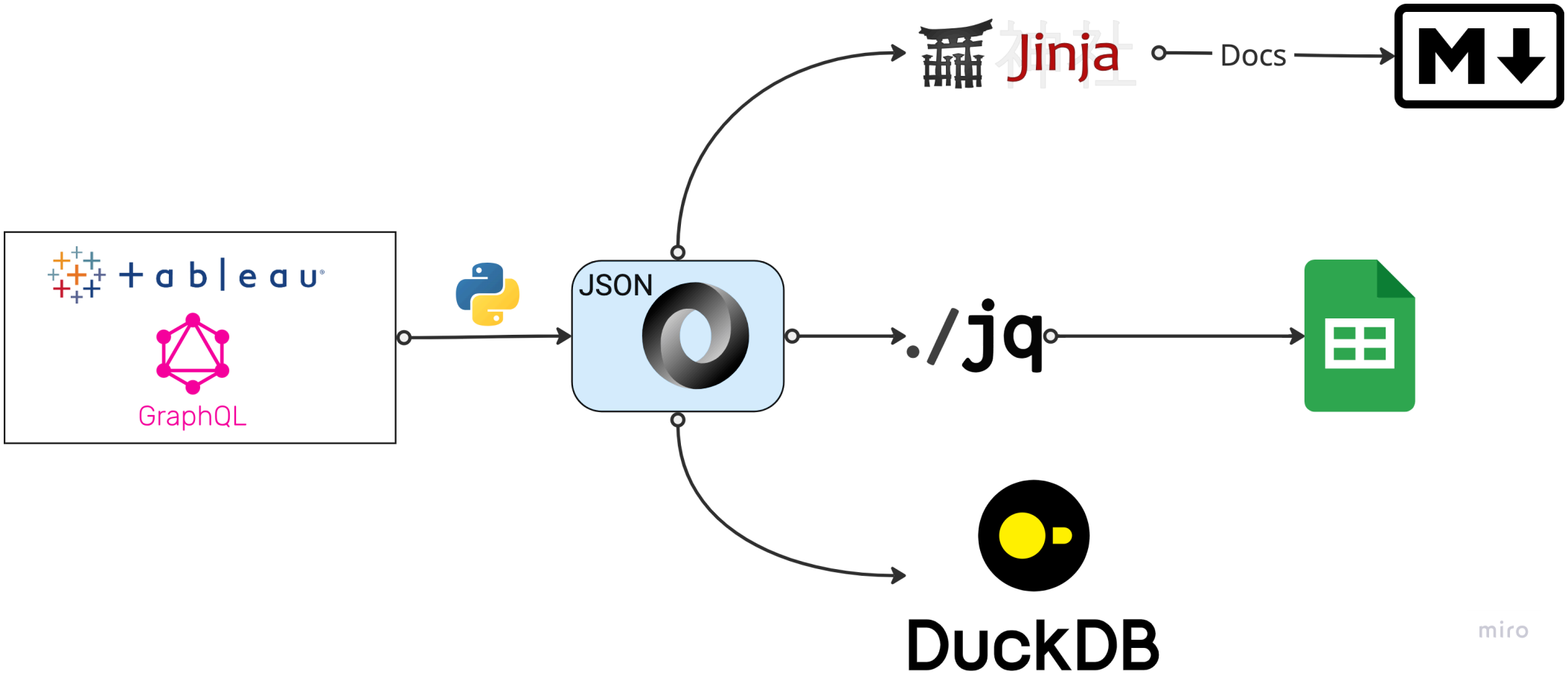
Just enough Python

```
with tableau_server.auth.sign_in(tableau_auth):
    variables = {"projectName": os.environ.get('PROJECT_NAME')}
    query = open('resources/graphql/simple_lineage.graphql').read()

    ## This is the important bit below
    query_results = tableau_server.metadata.query(query, variables=variables)
    with open('simple_lineage.json', 'w') as f:
        json.dump(query_results, f)
```

See: Tableau Server Python Client

(Meta)Data Engineering



EQUAL EXPERTS

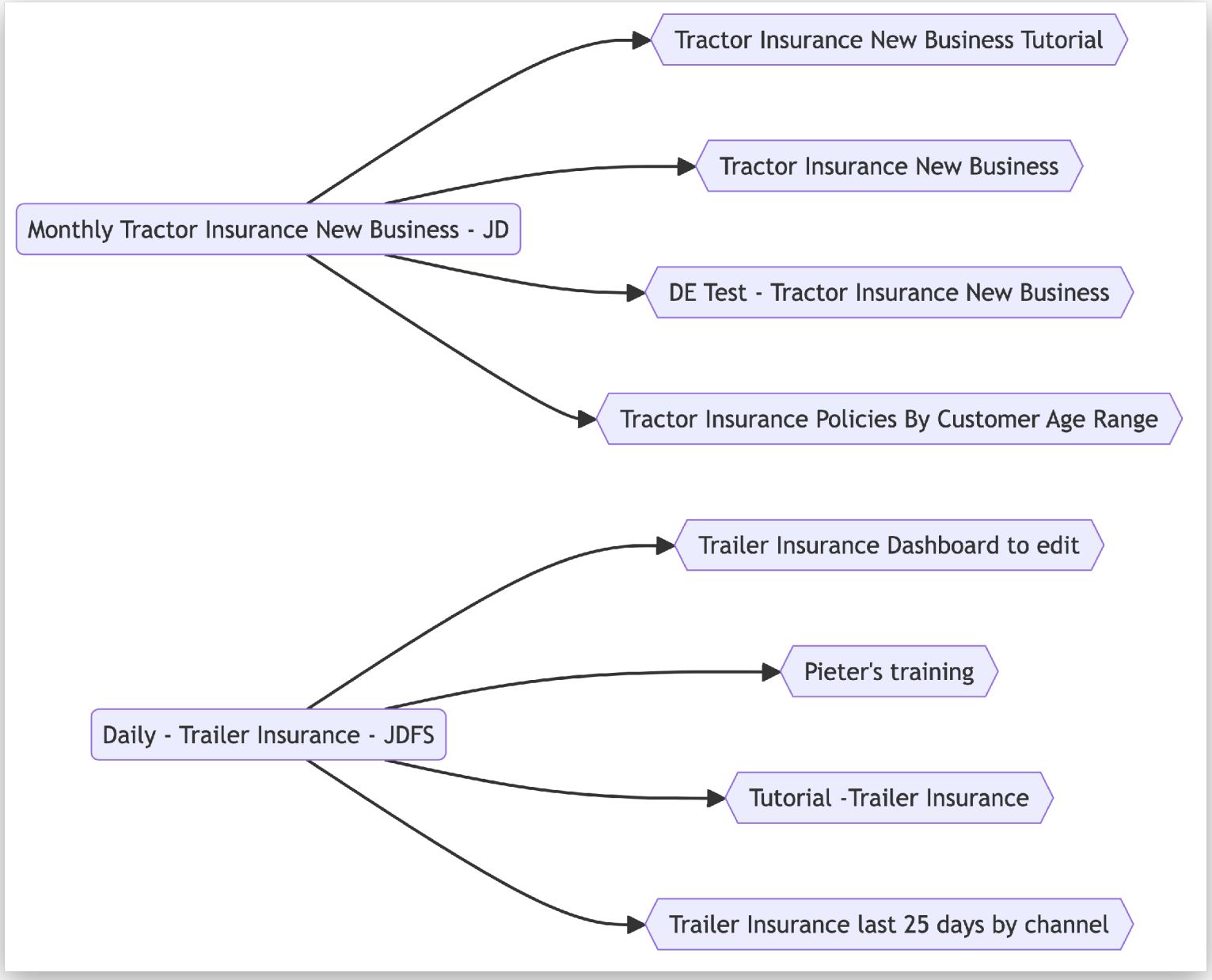
Jinja template 🥷

```
flowchart LR
{% for ds in data.publishedDatasources %}
    {{ ds.luid }}({{ ds.name }})
    {% for wb in ds.downstreamWorkbooks %}
      {% if wb.name != None %}
          {{ wb.luid }}["{{  wb.name }}"]
          {{ ds.luid }} --> {{ wb.luid }}
      {% endif %}
    {%  endfor %}
{%  endfor %}
```

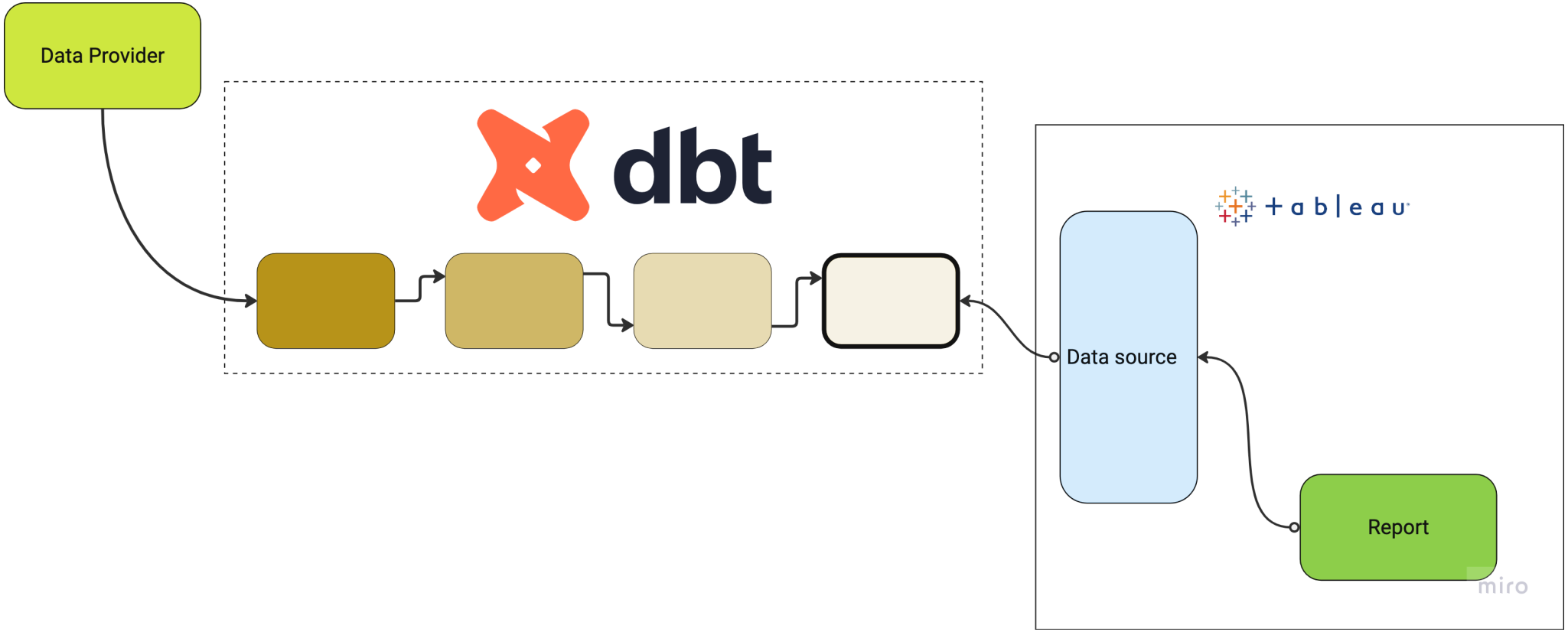
See: Jinja, MermaidJS flowcharts

EQUAL EXPERTS

The little mermaid 🧜



Extend right: data pipelines with dbt



EQUAL EXPERTS

When Business complains 😱

The figures are incorrect!

Data could be late?

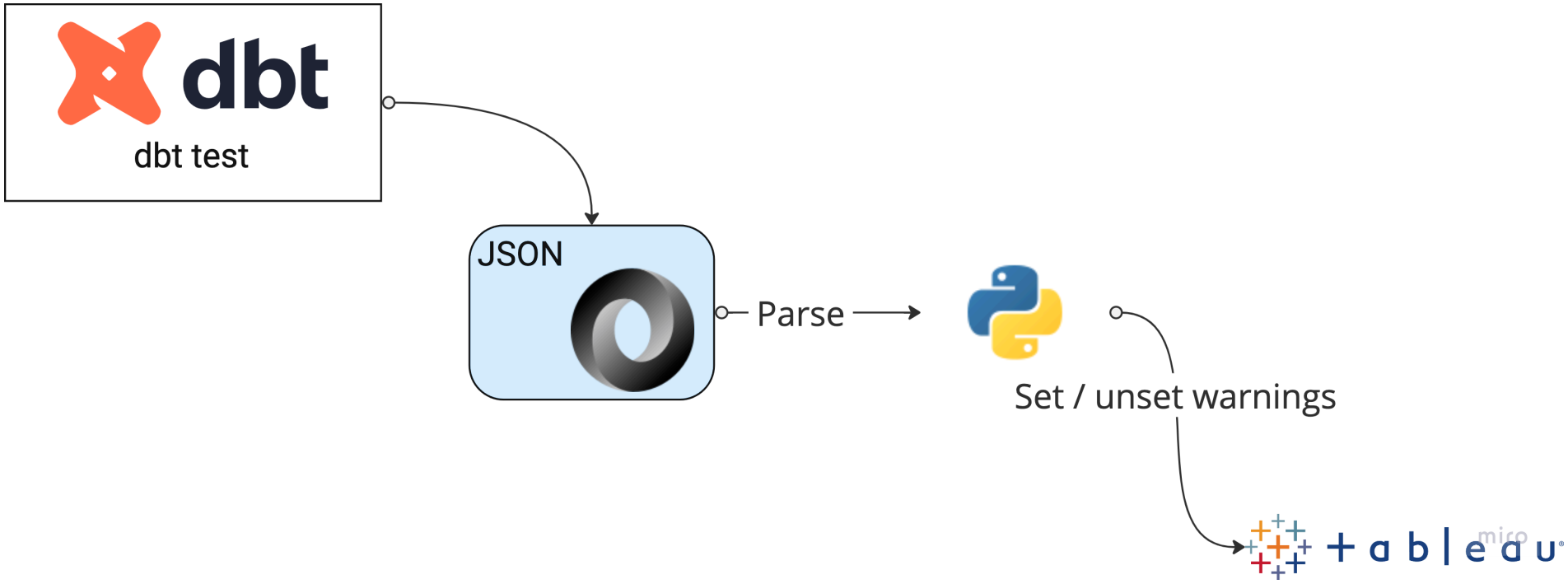
Data could be wrong?

EQUAL EXPERTS

Operability

- Data quality checks (e.g. dbt tests)
- Slack alerts on failures
- Not so business friendly 😢

From checks to labels



EQUAL EXPERTS

Sample dbt code

```
---
models:
  - name: tractor_sales
    columns:
      - name: created_at
        tests:
          - dbt_expectations.expect_row_values_to_have_recent_data:
              datepart: day
              interval: 1
              tags: ['tableau_recency']
```

See: dbt expectations

EQUAL EXPERTS

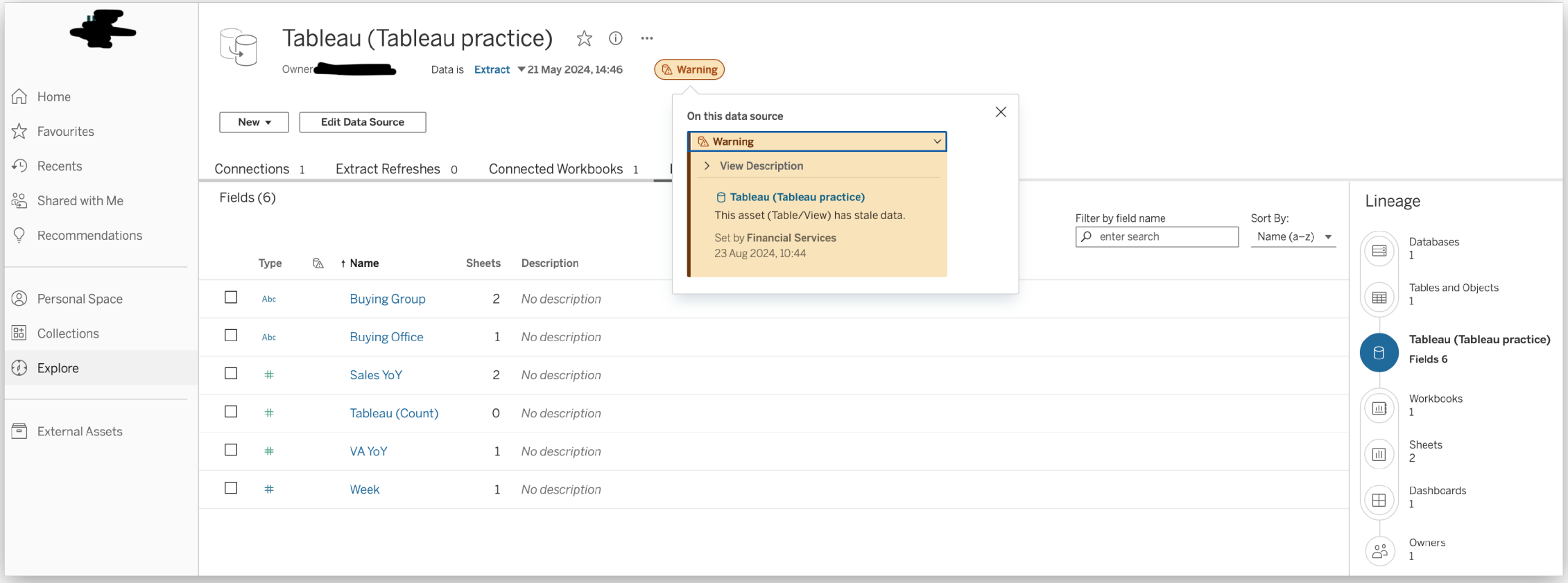
Just enough Python

```
table_item: TableItem = tableau_server.tables.get_by_id(luid)

data_quality_warning: DQWItem = DQWItem(
    warning_type=DQWItem.WarningType.WARNING,
    message="This asset (Table/View) has stale data.",
    active=True,
    severe=True)

tableau_server.tables.add_dqw(table_item, data_quality_warning)
```

What does it look like?



EQUAL EXPERTS

Shift left: should you push this to production?

```
SELECT applicationId AS application_id,
    datecreated as DATE_CREATED
from transactions;
```

EQUAL EXPERTS

SQLFluff says NO!!!

```
== [/Users/guillaumebelrose/Work/Presentations/dataconf24/dataconf-demo/resources/sql/lint.sql] FAIL
L:   1 | P:   1 | LT09 | Select targets should be on a new line unless there is
                       | only one select target. [layout.select_targets]
L:   1 | P:   7 | LT02 | Expected line break and indent of 4 spaces before
                       | 'applicationId'. [layout.indent]
L:   1 | P:   8 | CP02 | Unquoted identifiers must be consistently upper case.
                       | [capitalisation.identifiers]
L:   2 | P:  17 | CP01 | Keywords must be consistently upper case.
                       | [capitalisation.keywords]
L:   7 | P:   1 | LT12 | Files must end with a single trailing newline.
                       | [layout.end_of_file]
All Finished 📜 🎉!
```

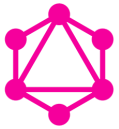
EQUAL EXPERTS

Keeping reporting assets consistent

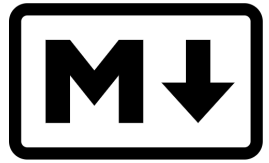
- Naming
  - Names | Descriptions
  - Tags
- Configuration
  - Caching
  - Database connection settings
- **Automated checks** over written guidelines

EQUAL EXPERTS

Roll Your Own Linter



GraphQL



Jinja

EQUAL EXPERTS

Shift left: The effects of UI-driven work

Time consuming to create | re-create

Time consuming to verify

Lack of audit trail

EQUAL EXPERTS

Lack of release pipeline

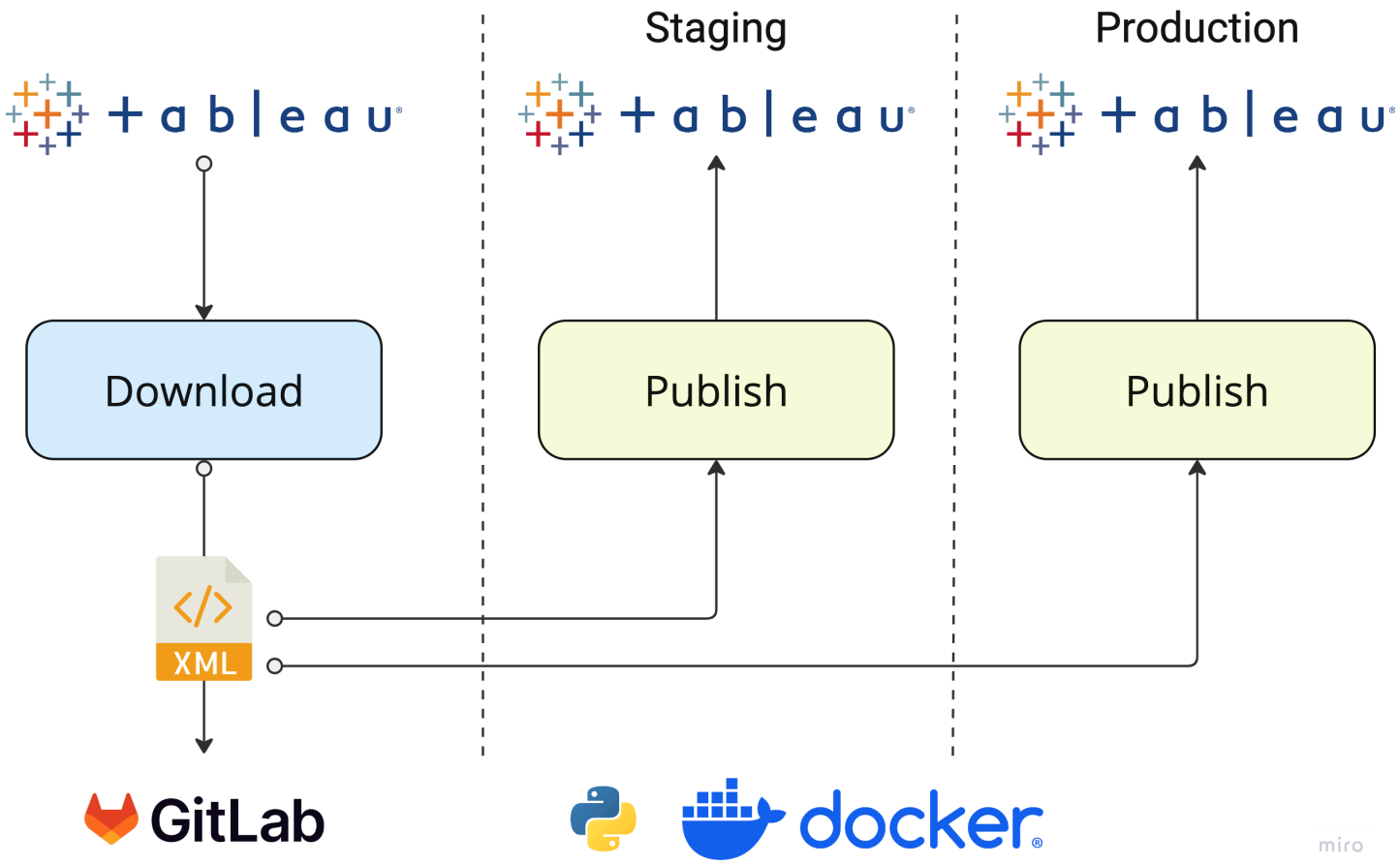
Less appreciation for segregated environments

Dev work becomes productionised

Self-serve vs dependable data

EQUAL EXPERTS

Data Source As Code



EQUAL EXPERTS

What it enables

Improved quality: dev -> qa -> prod

Common deployment patterns for reporting and data pipelines

Audit trail

EQUAL EXPERTS

To recap

EQUAL EXPERTS

Tales from production

- Applied this thinking to our domain
- Easier to adopt for green field projects
- Many areas are ripe for automation
- **Ways of working** just as crucial

EQUAL EXPERTS

Thank you!

To my awesome team

...and the awesome audience 😄

PS: don't ask me about PowerBI