How Airflow solves the coordination of decentralised teams at Vinted

3.0

Oscar Ligthart
Rodrigo Loredo

Vinted

Make second-hand first choice worldwide

Millions
of registered members

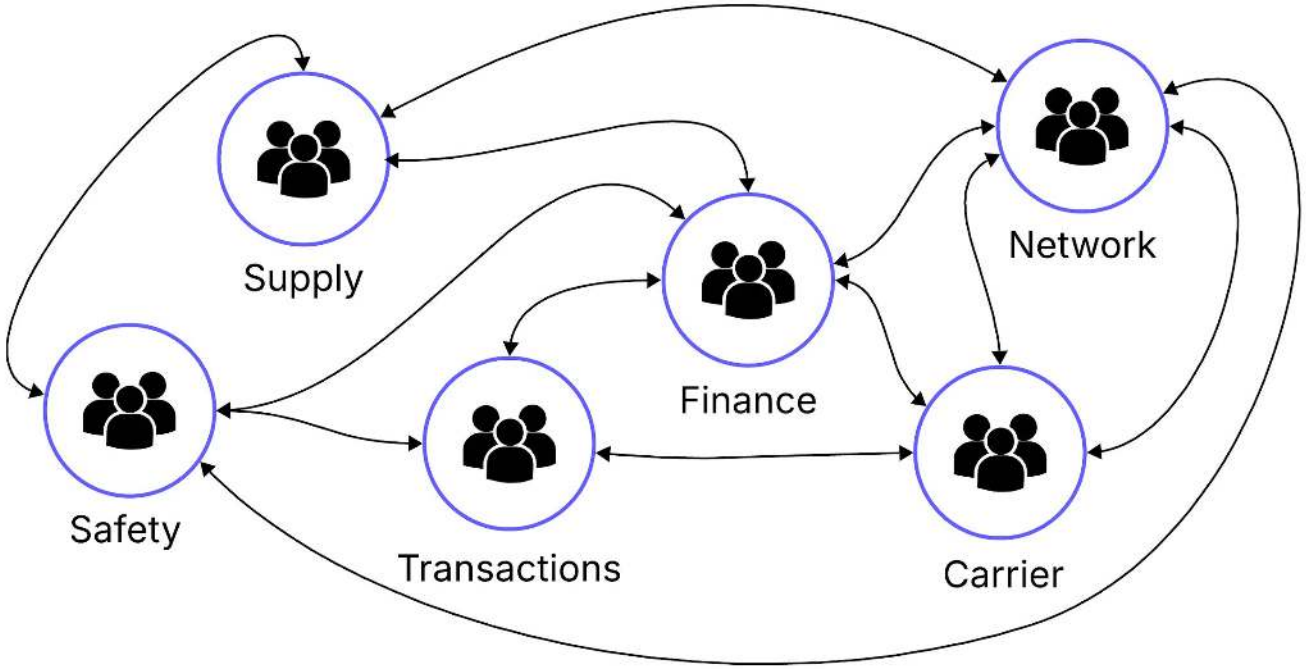
20+
countries

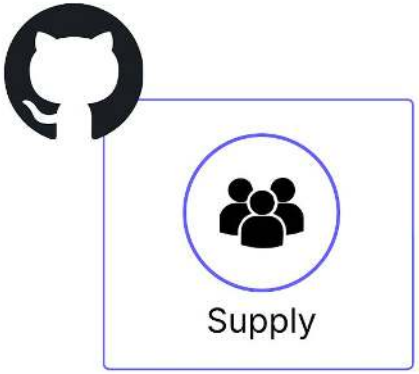
2000+
talents

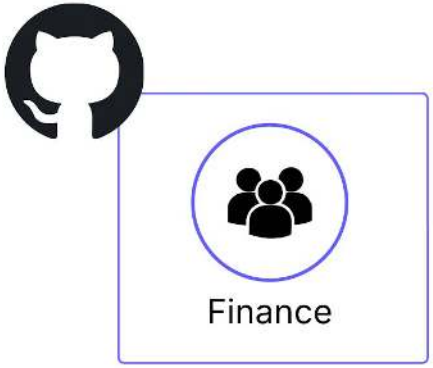
10
workplaces

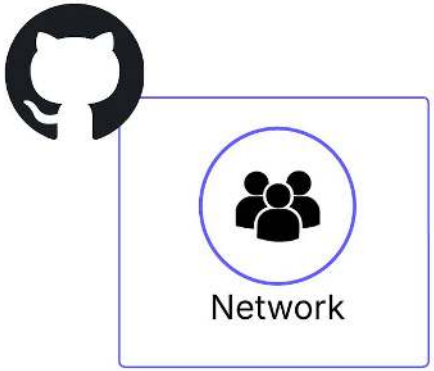
2000+

talents

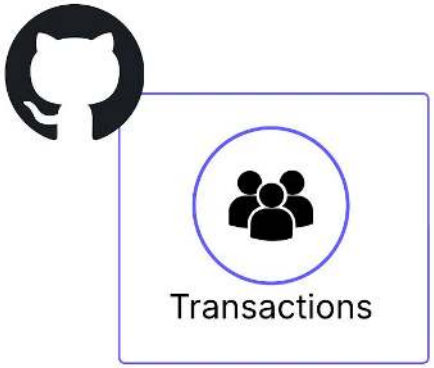


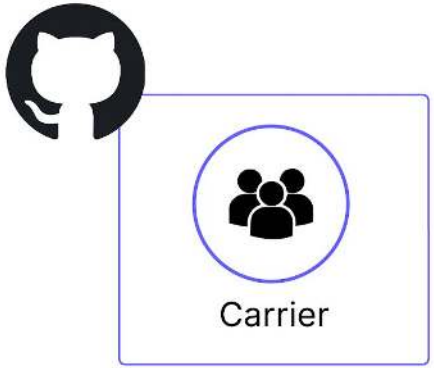

Supply

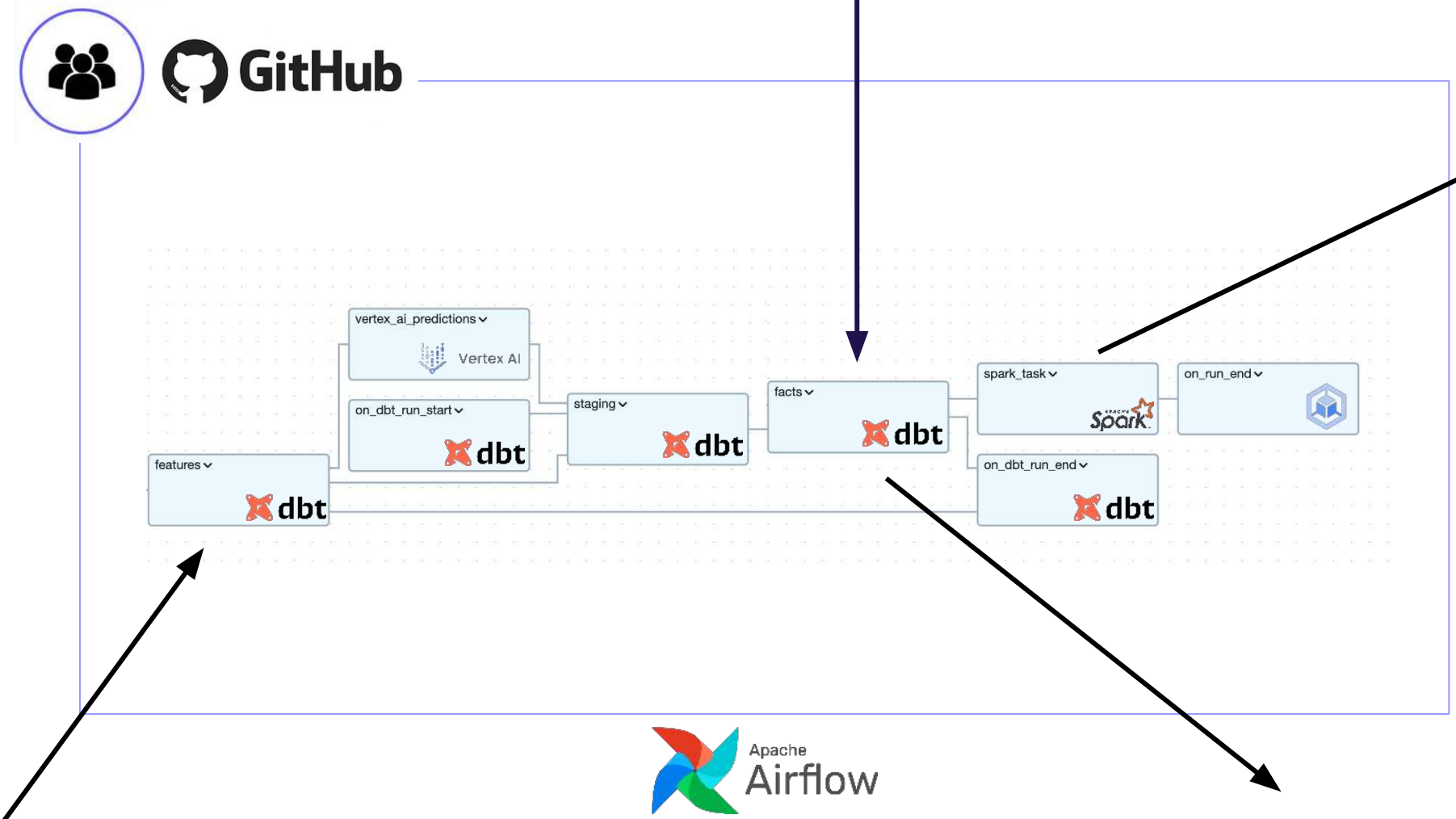

Finance

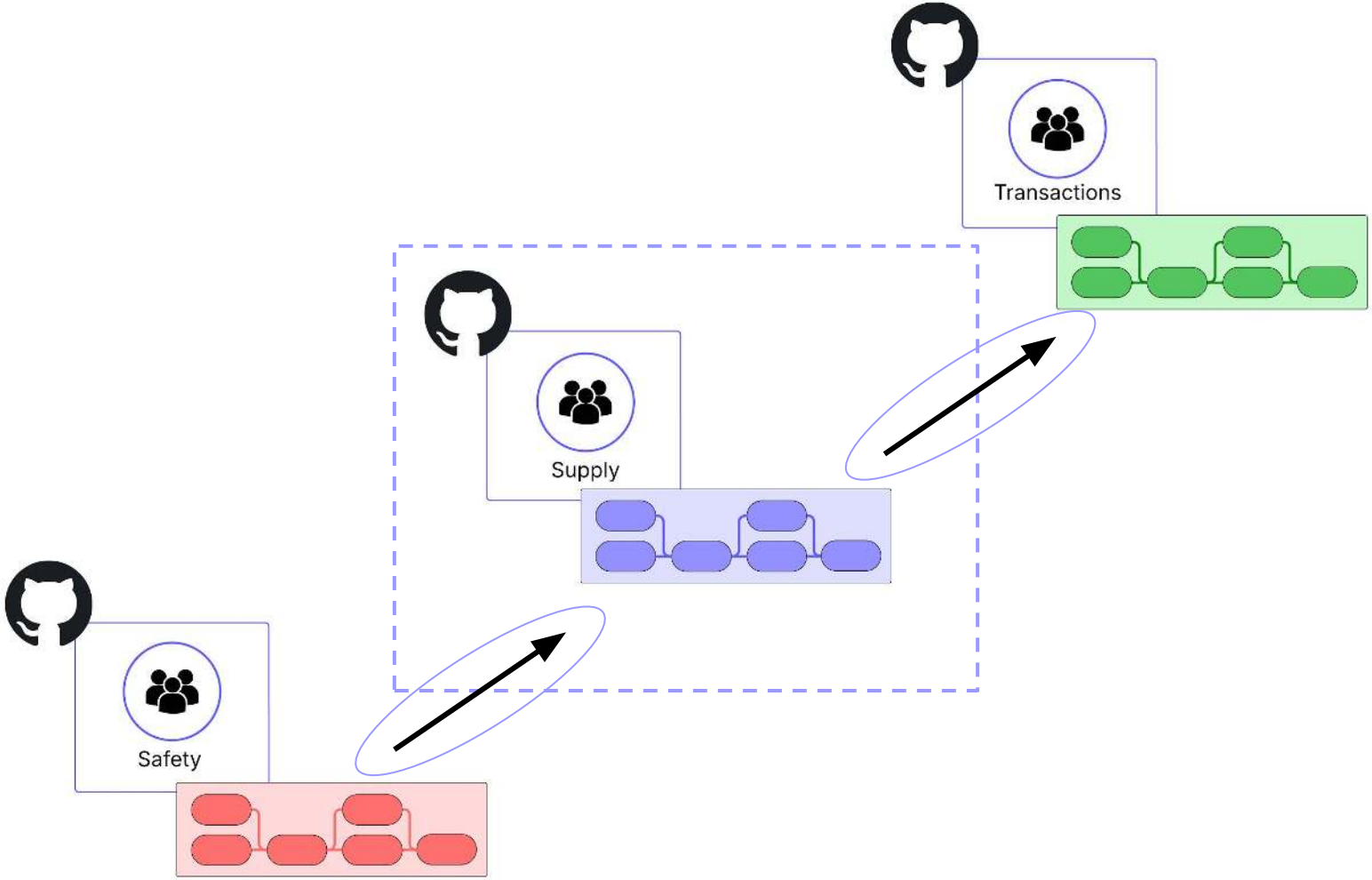

Network

Safety

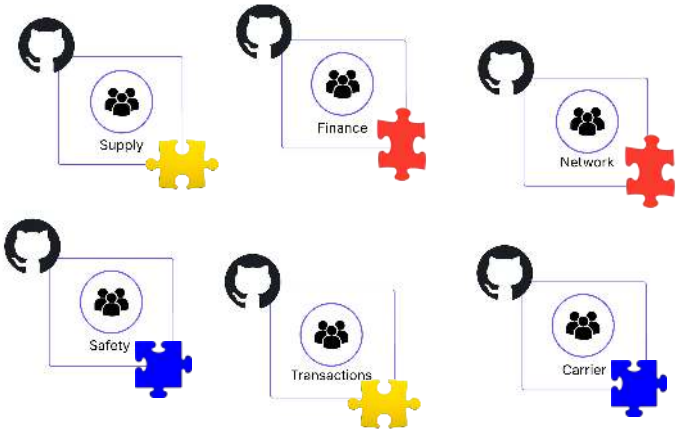

Transactions


Carrier


GitHub
vertex_ai_predictions
Vertex AI
on_dbt_run_start
dbt
features
dbt
staging
dbt
facts
dbt
spark_task
Spark
on_run_end
on_dbt_run_end
dbt
Apache
Airflow

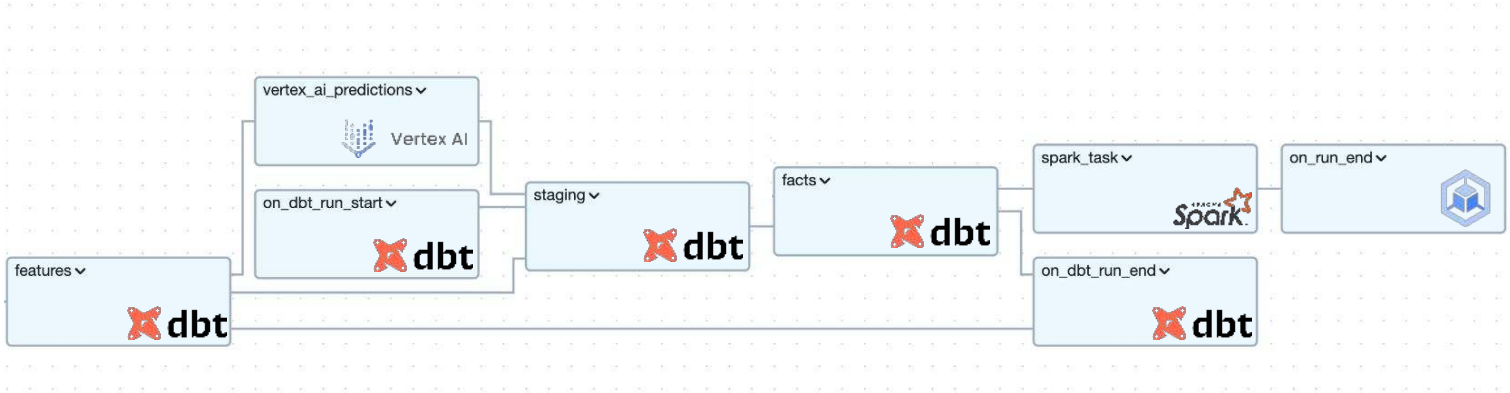

Transactions
Supply
Safety

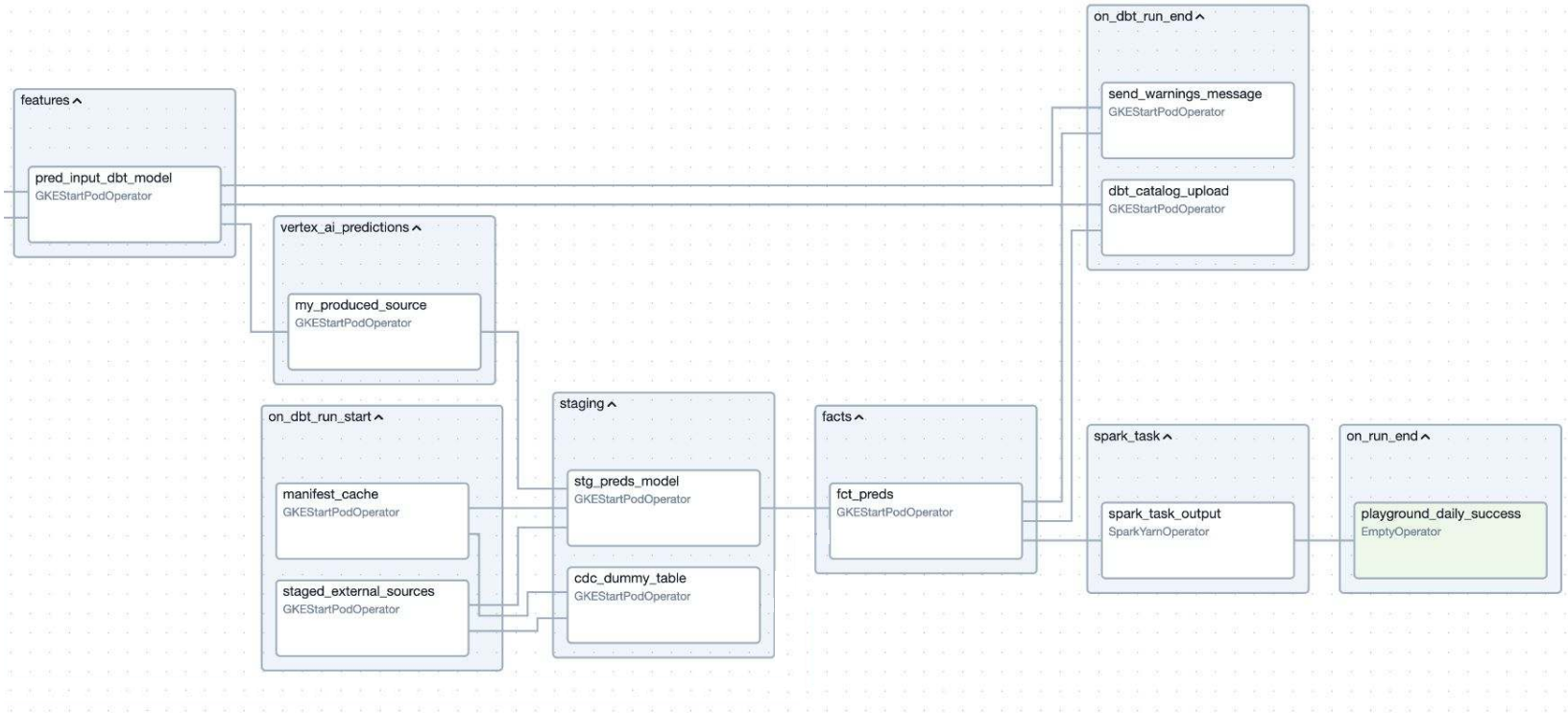
OUR DATA TEAM IS A BOTTLENECK
DECENTRALIZE!
IT'S TOO HARD TO COORDINATE
CENTRALIZE!
IT'S NOT SCALABLE
DECENTRALIZE A LITTLE!
imgflip.com
MemeCenter


Supply
Finance
Network
Safety
Transactions
Carrier

PUZZLE
150+
DAGs
Vinted
20k+
tasks/day
Airflow

Generating DAGs


vertex_ai_predictions
Vertex AI
on_dbt_run_start
dbt
features
dbt
staging
dbt
facts
dbt
spark_task
Spark
on_run_end
on_dbt_run_end
dbt


features
pred_input_dbt_model
GKEStartPodOperator
vertex_ai_predictions
my_produced_source
GKEStartPodOperator
on_dbt_run_start
manifest_cache
GKEStartPodOperator
staged_external_sources
GKEStartPodOperator
staging
stg_preds_model
GKEStartPodOperator
cdc_dummy_table
GKEStartPodOperator
facts
fct_preds
GKEStartPodOperator
on_dbt_run_end
send_warnings_message
GKEStartPodOperator
dbt_catalog_upload
GKEStartPodOperator
spark_task
spark_task_output
SparkYarnOperator
on_run_end
playground_daily_success
EmptyOperator

DAG generator design: User interface

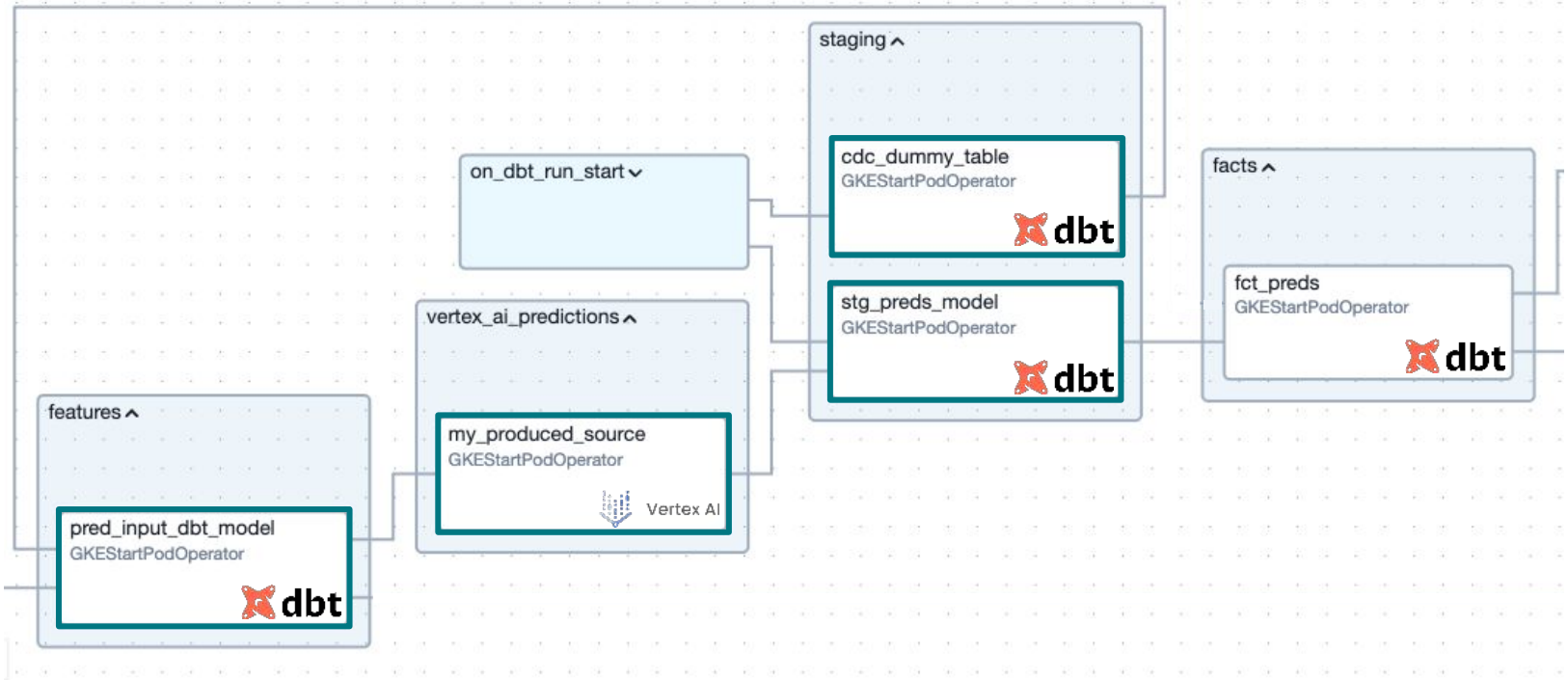
```yaml
version: 0.17
tags:
  - tag_1
  - long_tag

alerting:
  channel: "#some-alerts-channel"
  tag:
    - "@amazing.person" # a person
    - "!subteam^S05VAFP4A6N" # a Team

dbt_settings:
  enabled: true
  on_run_start:
    - stage_external_sources
  on_run_end:
    - send_warnings_message

tasks:
   - kind: universal
     group: some_task_group_name
     conditions:
      - some_source
     assets:
      - produced_output_table
     timeout:
       minutes: 30
     image: image_name:tag
     command: poetry run python -m my_module.main

exposed_assets:
  - regex: "mrt_daily_exposed"
    slo:
      daily:
        hours: 4
```


staging
on_dbt_run_start
cdc_dummy_table
GKEStartPodOperator
dbt
facts
vertex_ai_predictions
stg_preds_model
GKEStartPodOperator
dbt
fct_preds
GKEStartPodOperator
dbt
features
my_produced_source
GKEStartPodOperator
Vertex AI
pred_input_dbt_model
GKEStartPodOperator
dbt

Connecting Tasks

dbt_models.yml

```yaml
models:
  - name: input_dbt_model
    config:
      meta:
        scheduler:
          pipelines:
            - hourly
```

scheduler.yaml

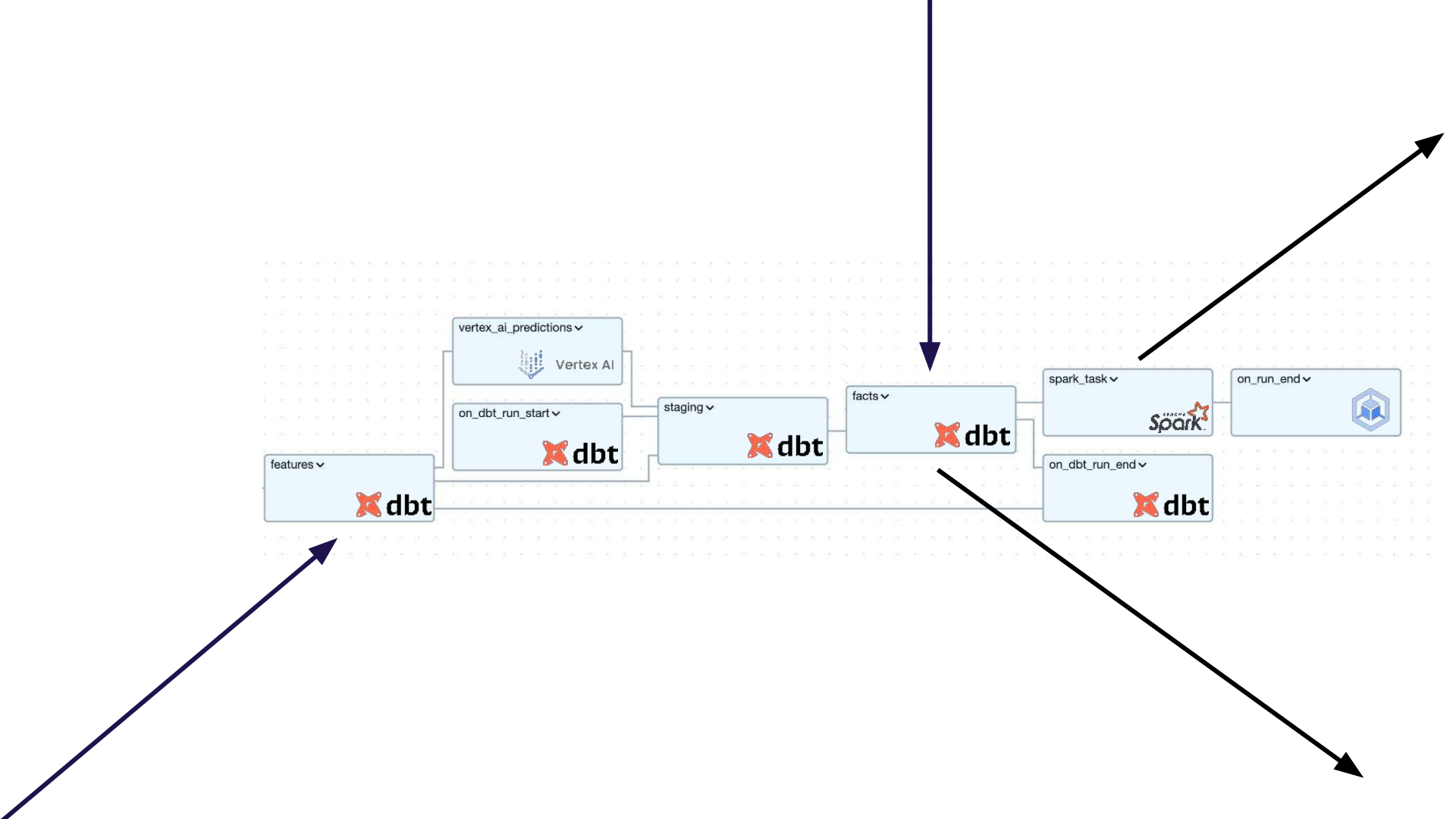
```yaml
tasks:
   - kind: vertex-ai
     group: predictions
     conditions:
       - input_dbt_model
     assets:
        - my_produced_source
```

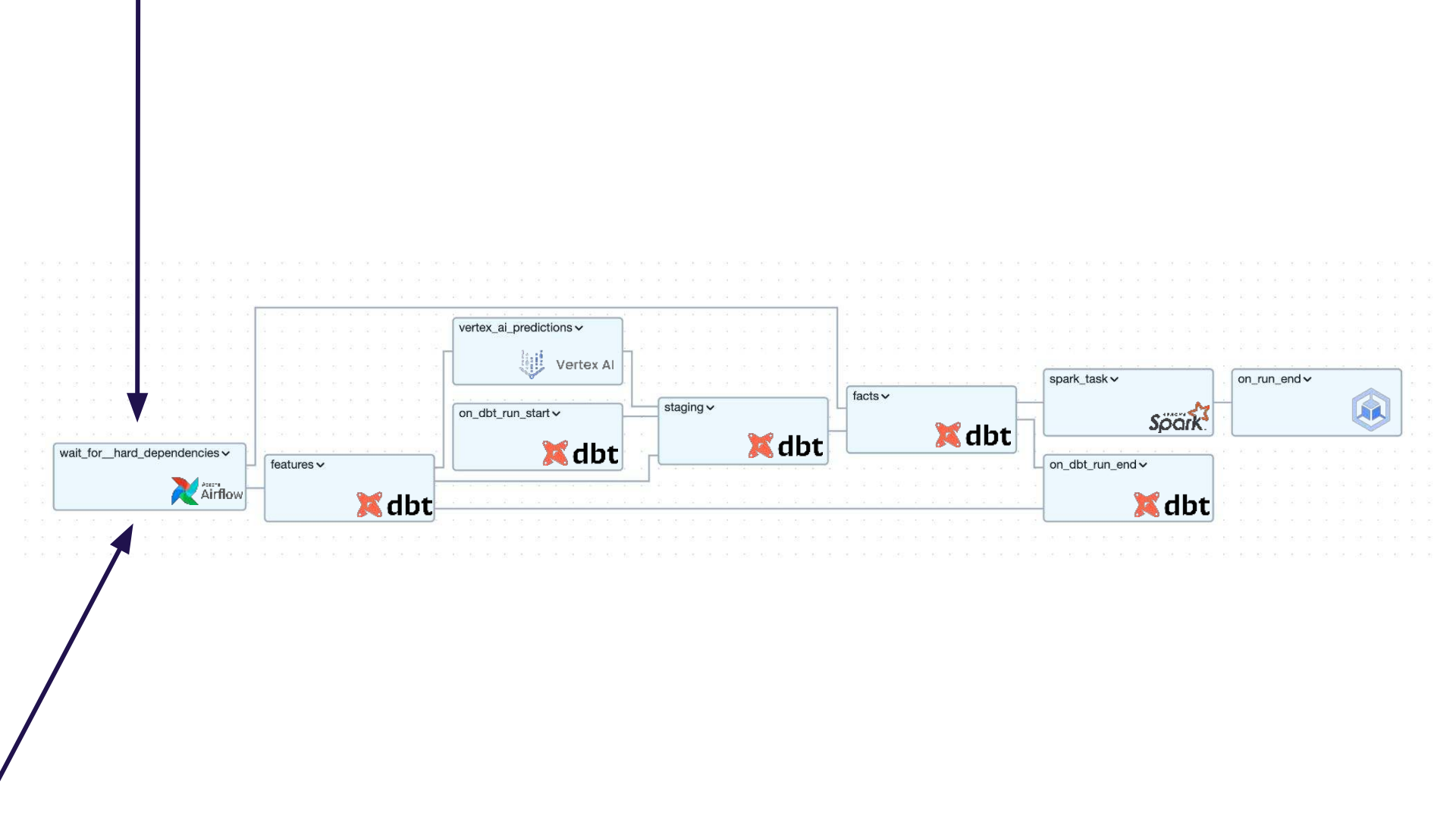
predictions_sources.yml

```yaml
sources:
  - name: predictions
    tables:
      - name: my_produced_source
        columns:
        ...
```

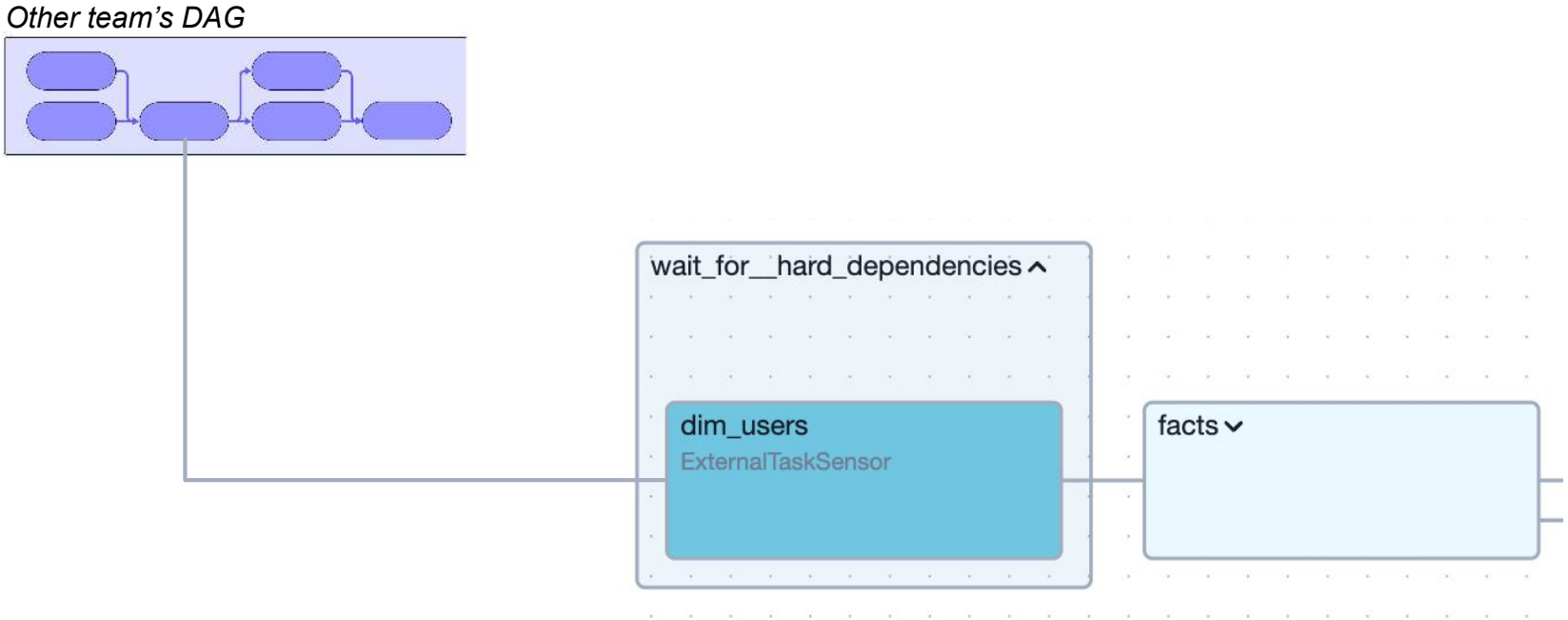
some_model.sql

```sql
...
select customer_id, predicted_value
from {{ source("predictions", "my_produced_source") }}
...
```


vertex_ai_predictions
Vertex AI
on_dbt_run_start
dbt
features
dbt
staging
dbt
facts
dbt
spark_task
Spark
on_run_end
on_dbt_run_end
dbt


wait_for__hard_dependencies
Airflow
features
dbt
vertex_ai_predictions
Vertex AI
on_dbt_run_start
dbt
staging
dbt
facts
dbt
spark_task
Spark
on_dbt_run_end
dbt
on_run_end

Using other team's assets



Using other team's assets

scheduler.yaml

```yaml
exposed_assets:
  # All dimensions
  - regex: "dim_.*"
    slo:
      daily:
        hours: 6
```

Asset registry object

```
{
  "some_team": {
    "produced": {
      "dim_customers": {
        "dag_prefix": "some_team",
        "task_group": "dimensions",
        "task_id": "dim_customers",
        "pipeline": "daily",
        "slo_seconds": 21600
      }
    }
}
```

CI/CD validation

```yaml
sources:
  - name: global_assets
    tables:
      - name: dim_customers
```

```
...
select customer_id, country_id
from {{ source("predictions", "dim_customers") }}
...
```

Developer Flow

playground/models/global_assets/global_sources.yml +4 Viewed

```
@@ -0,0 +1,4 @@
+ sources:
+   - name: shipping
+     tables:
+       - name: fct_shipments
```

playground/models/marts/my_new_model.sql +3 Viewed

```
@@ -0,0 +1,3 @@
+ select *
+ from {{ ref("fct_preds") }}
+ left join {{ source("shipping", "fct_shipments") }} as s on fct_preds.shipment_id = s.shipment_id
```

🗓️ Scheduler report: playground

✅ No breaking changes found in project registry.

Scheduler state diff

Dags: 0 added, 0 removed
Tasks: 1 added, 0 removed
Global dependencies: 1 added, 0 removed

Tasks and global dependencies

playground_daily

+ global_dependency.fct_shipments
+ marts.my_new_model

🗒️ Scheduler report: playground

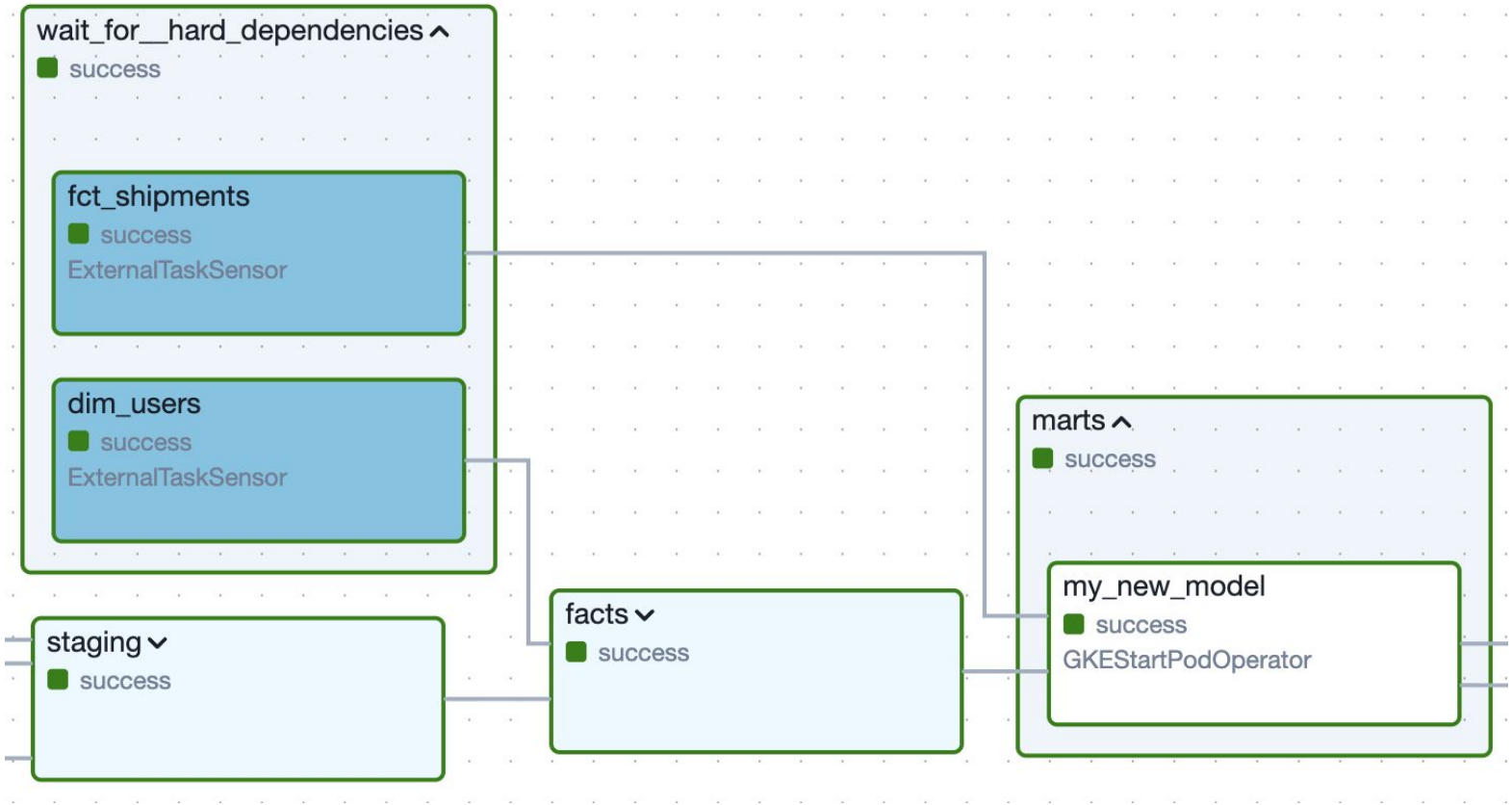
⚠️ Breaking changes found in project registry

Asset test_users

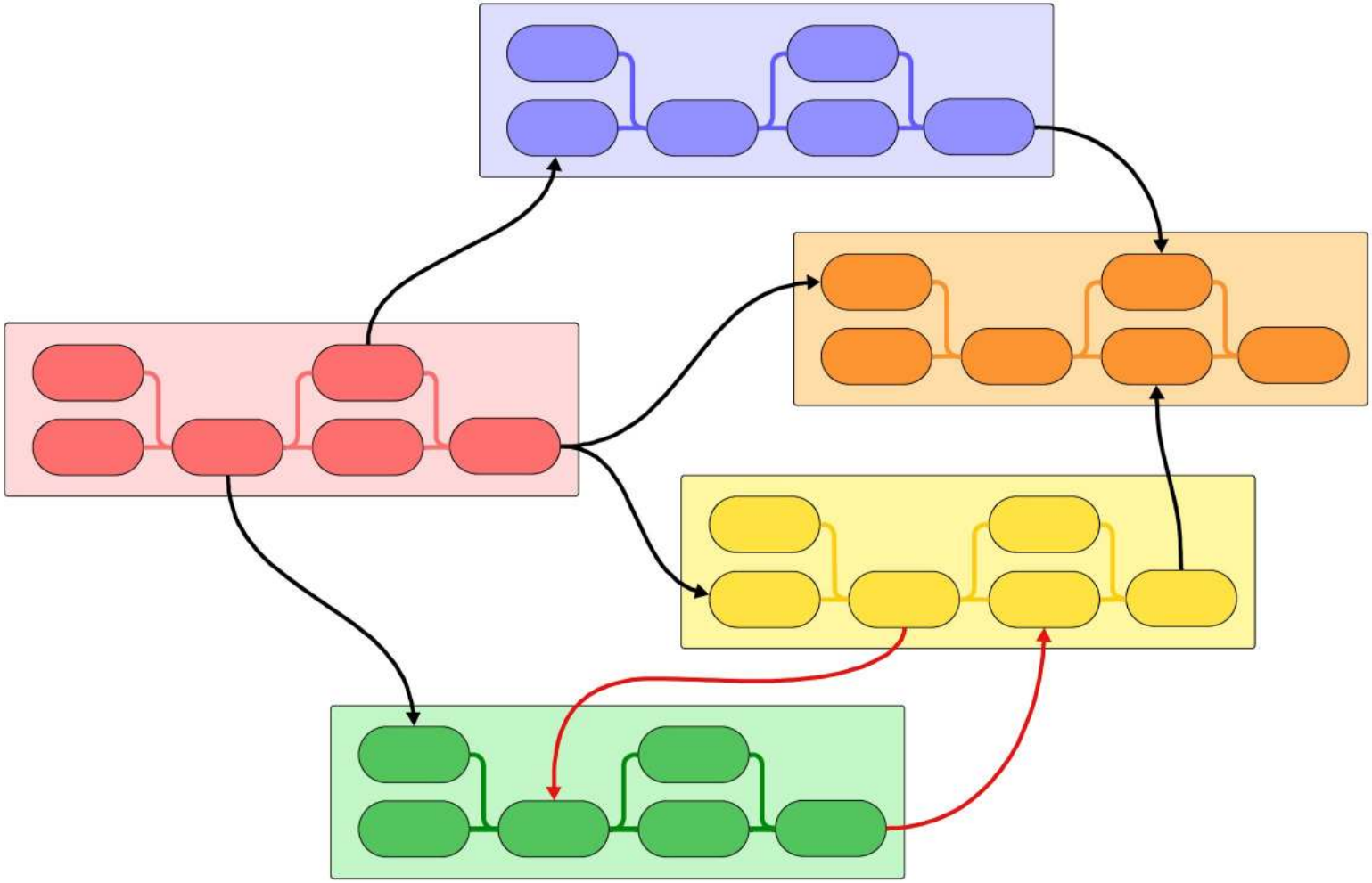
- Pipeline changed from hourly to daily.
- SLO increased from 2700 to 162000.

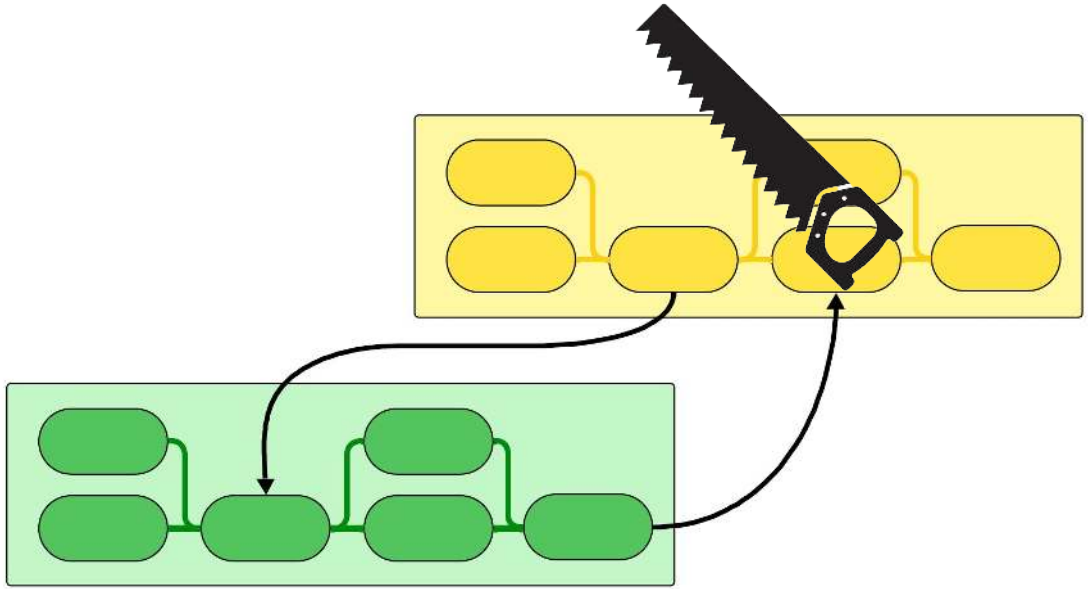
Consumers:

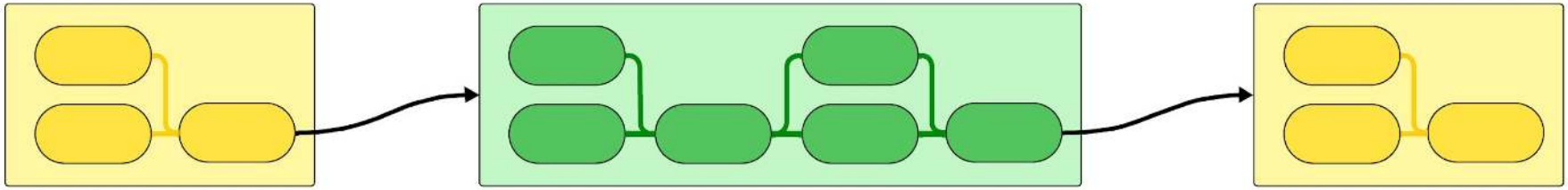
- *some_team*:#some-team-contact-channel

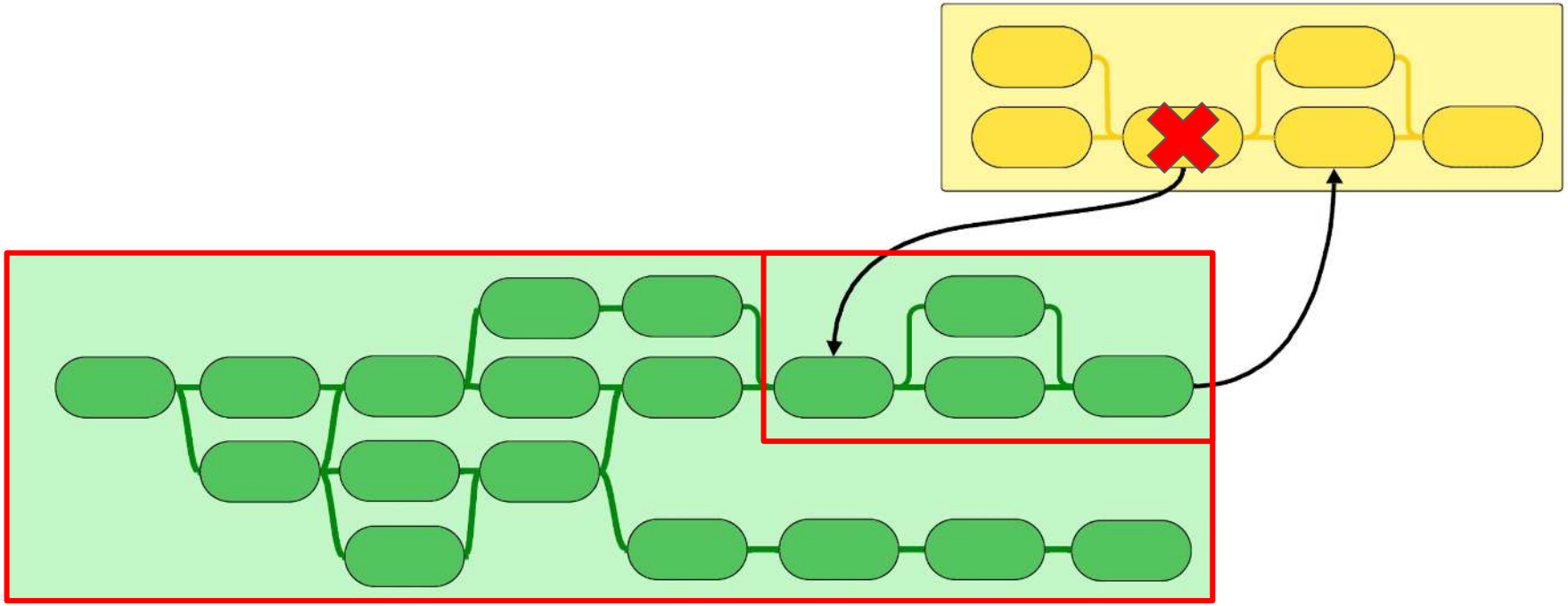

wait_for__hard_dependencies
success
fct_shipments
success
ExternalTaskSensor
dim_users
success
ExternalTaskSensor
staging
success
facts
success
marts
success
my_new_model
success
GKEStartPodOperator

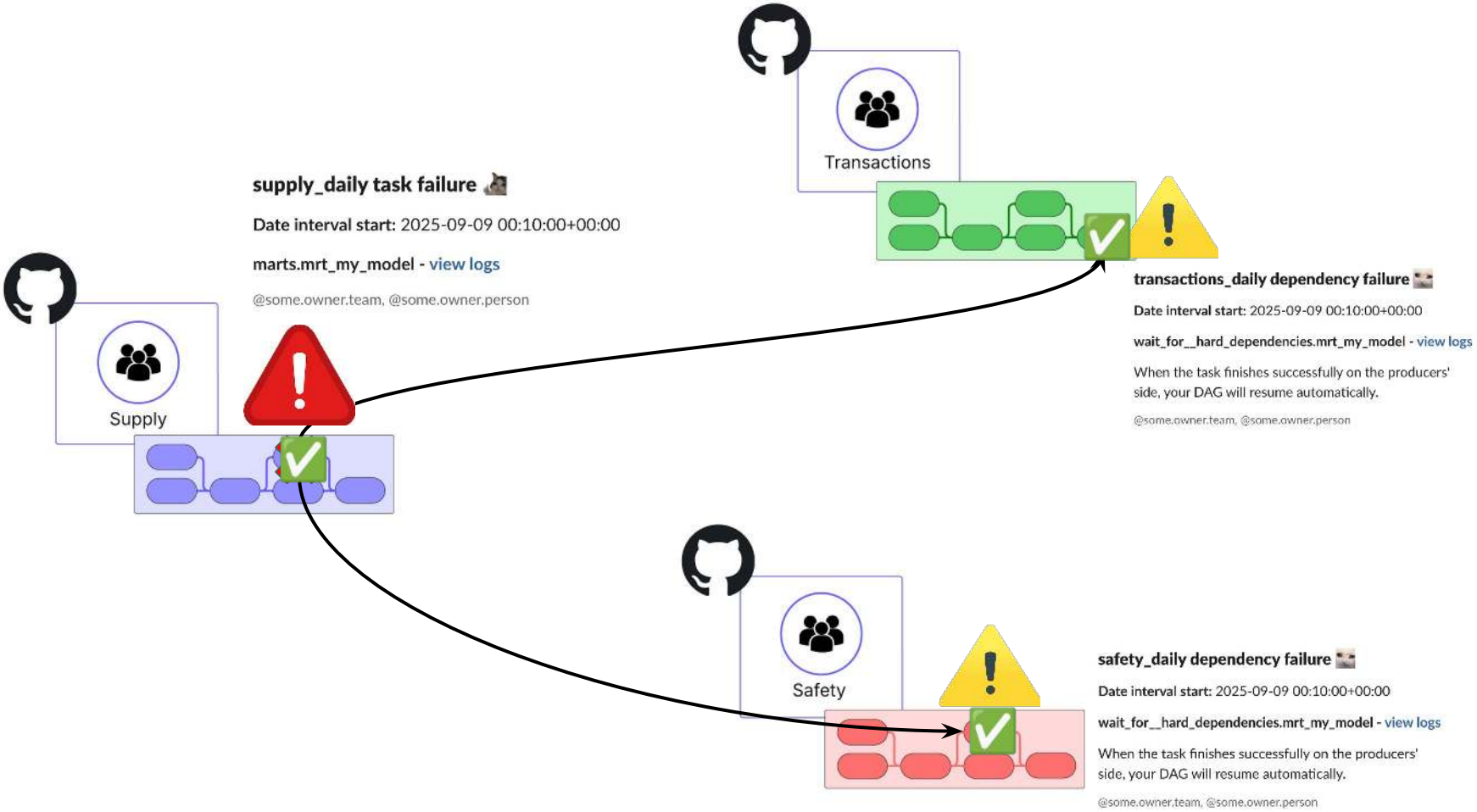
Sensor Setup





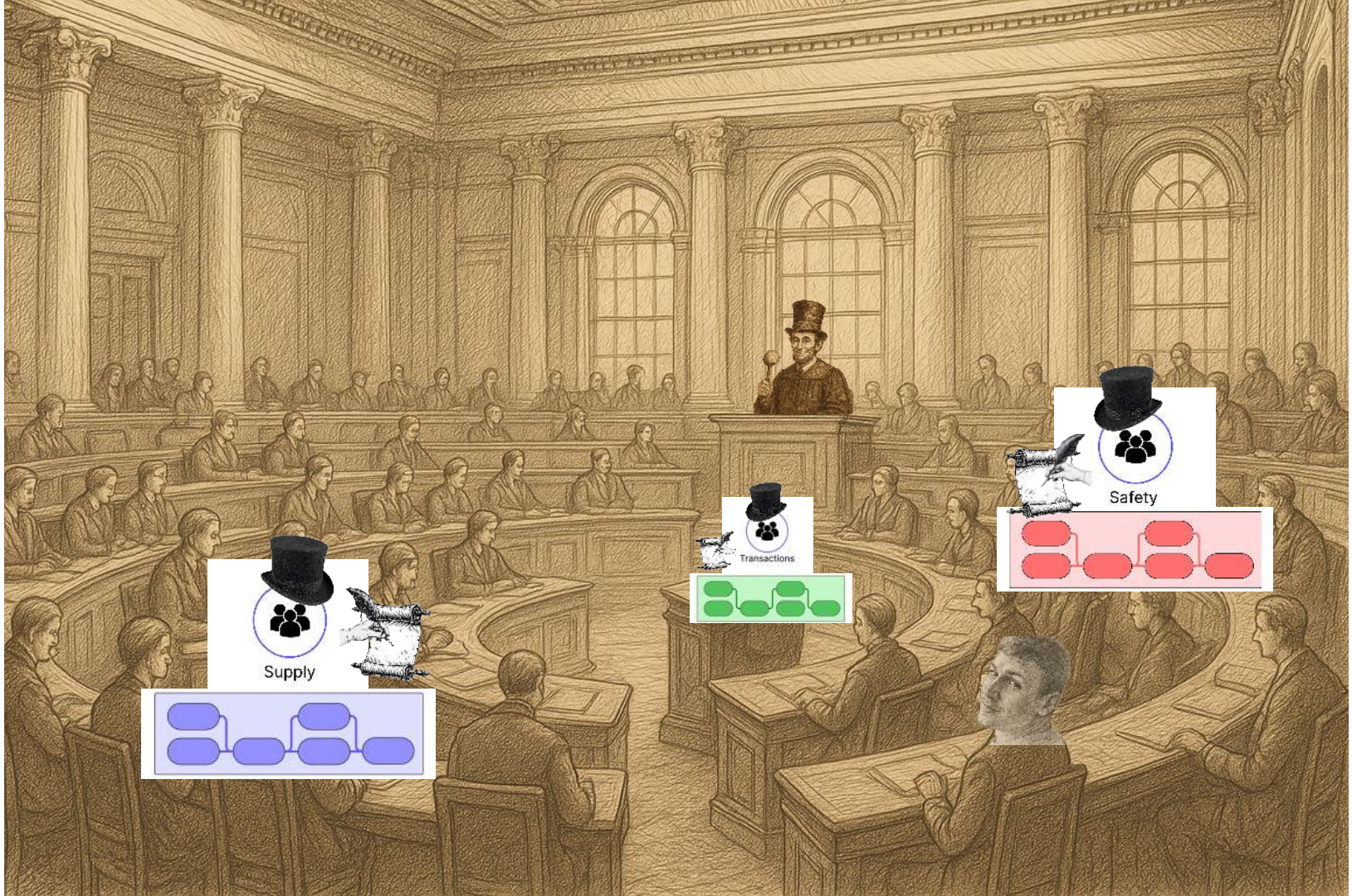





Transactions
supply_daily task failure
Date interval start: 2025-09-09 00:10:00+00:00
marts.mrt_my_model - view logs
@some.owner.team, @some.owner.person
transactions_daily dependency failure
Date interval start: 2025-09-09 00:10:00+00:00
wait_for__hard_dependencies.mrt_my_model - view logs
When the task finishes successfully on the producers' side, your DAG will resume automatically.
@some.owner.team, @some.owner.person
Supply
Safety
safety_daily dependency failure
Date interval start: 2025-09-09 00:10:00+00:00
wait_for__hard_dependencies.mrt_my_model - view logs
When the task finishes successfully on the producers' side, your DAG will resume automatically.
@some.owner.team, @some.owner.person

The New Normal

HOTEL
GENERAL STORE
Transactions
Supply
Safety


Supply
Transactions
Safety

WITH OUR
DAG GENERATOR
WITH DAGS BUILT BY
15 TEAMS AND LEGACY
CODE SUGGESTED BY AN LLM
UPGRADING
AIRFLOW
imgflip.com

Thank you!

Questions?