



How Airflow solves the coordination of decentralised teams at Vinted

Oscar Lighthart
Rodrigo Loredo

3.0

Vinted

Make second-hand first
choice worldwide



Millions

of registered members

2000+

talents

20+

countries

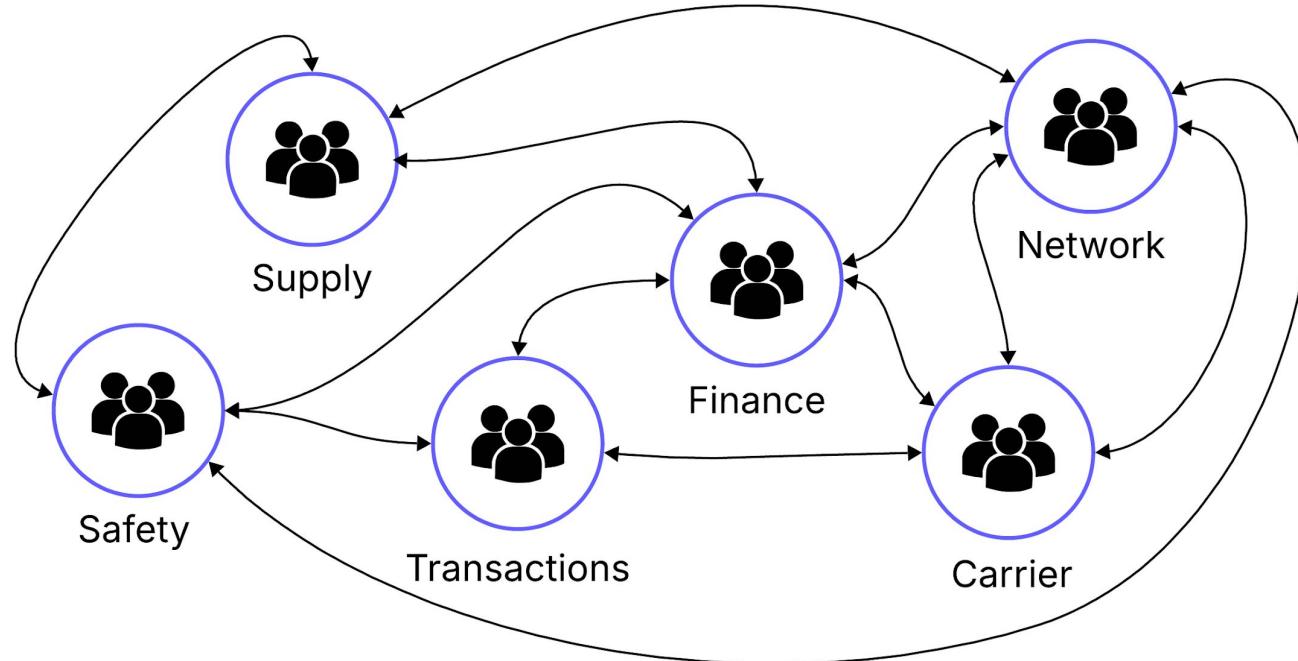
10

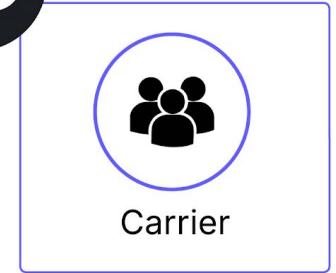
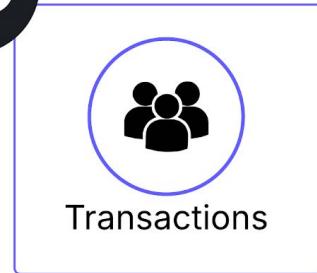
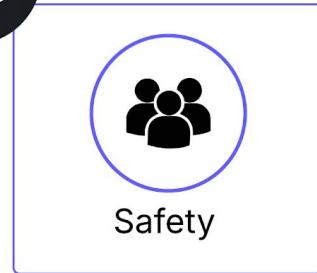
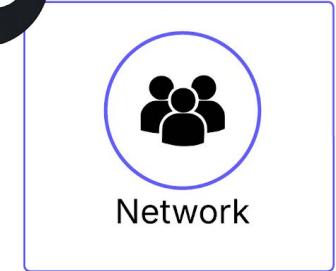
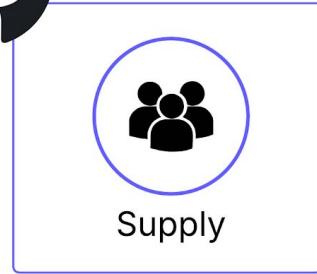
workplaces

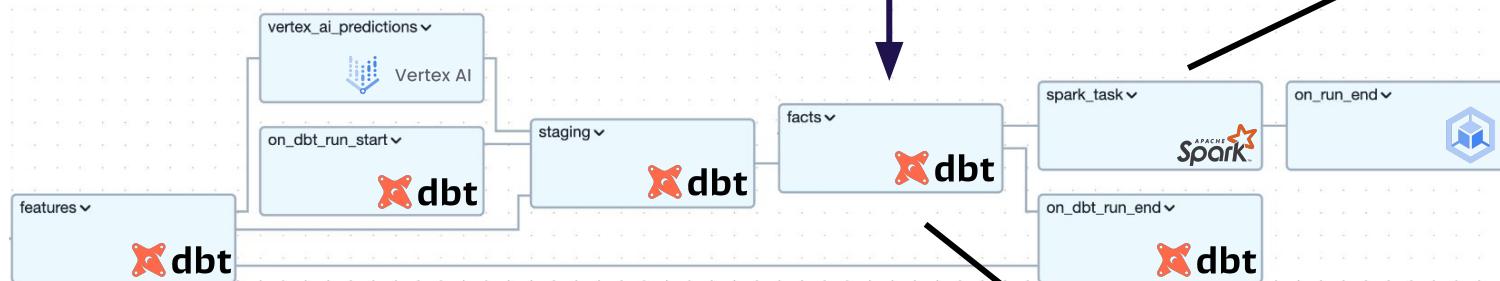


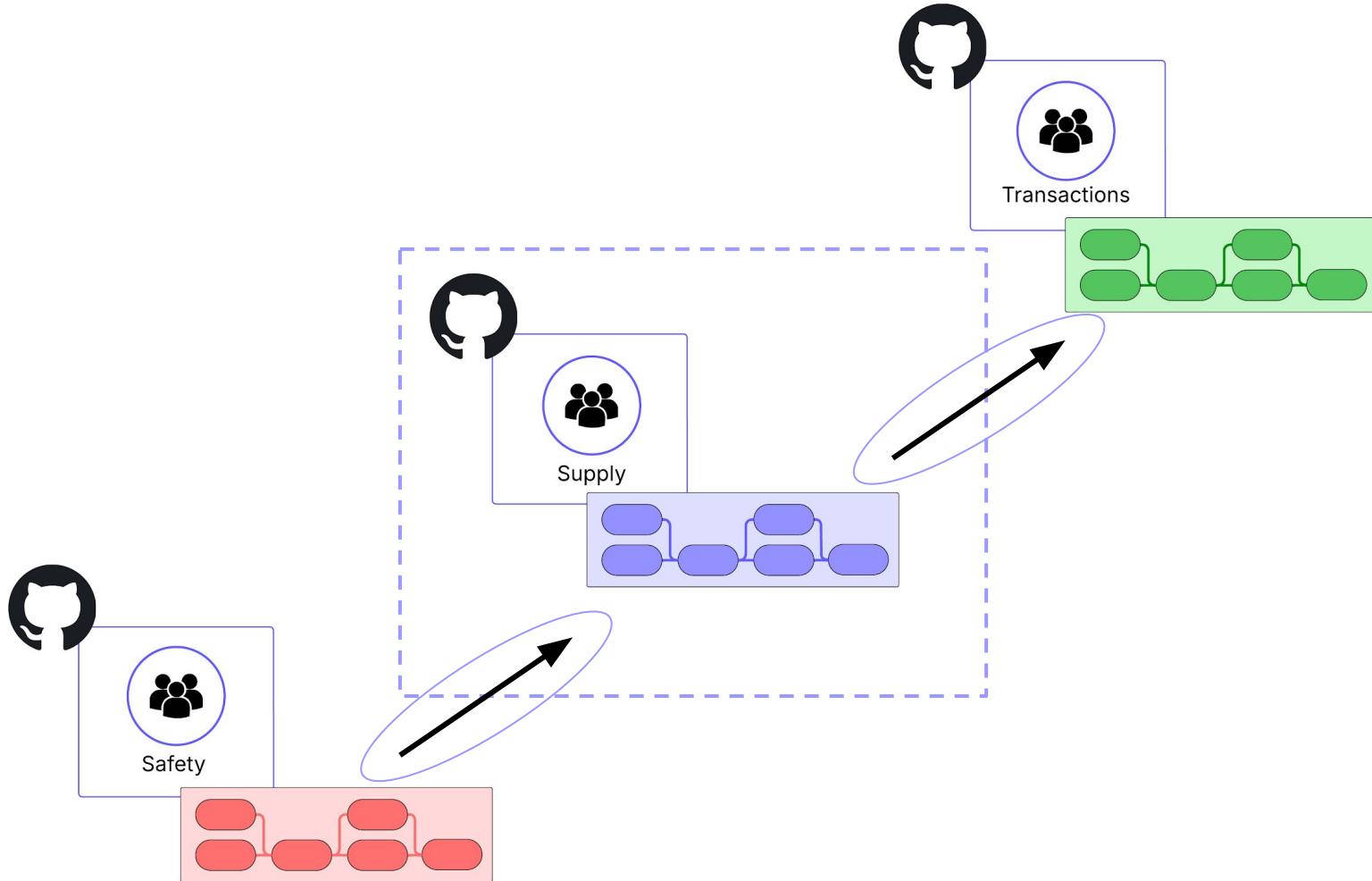
2000+

talents

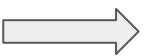
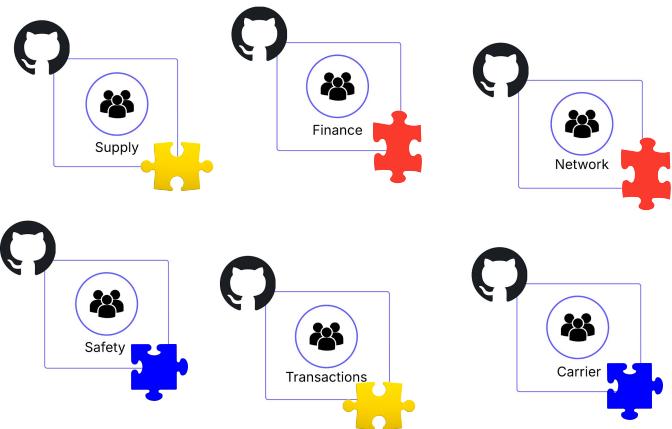




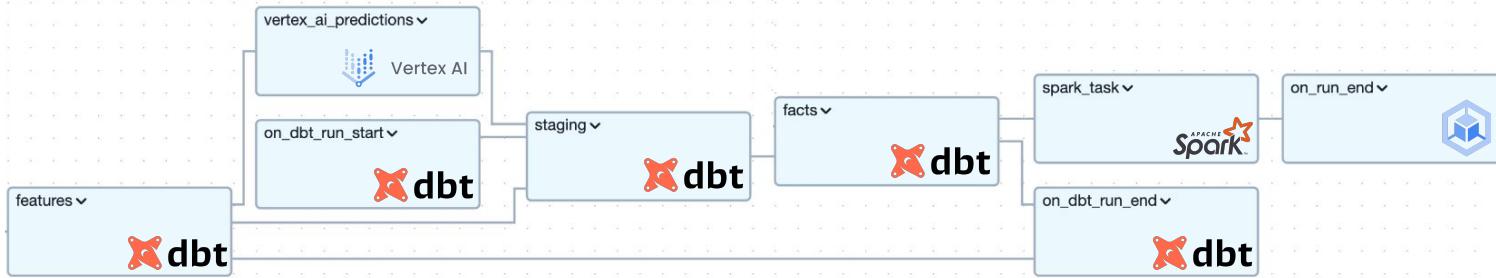


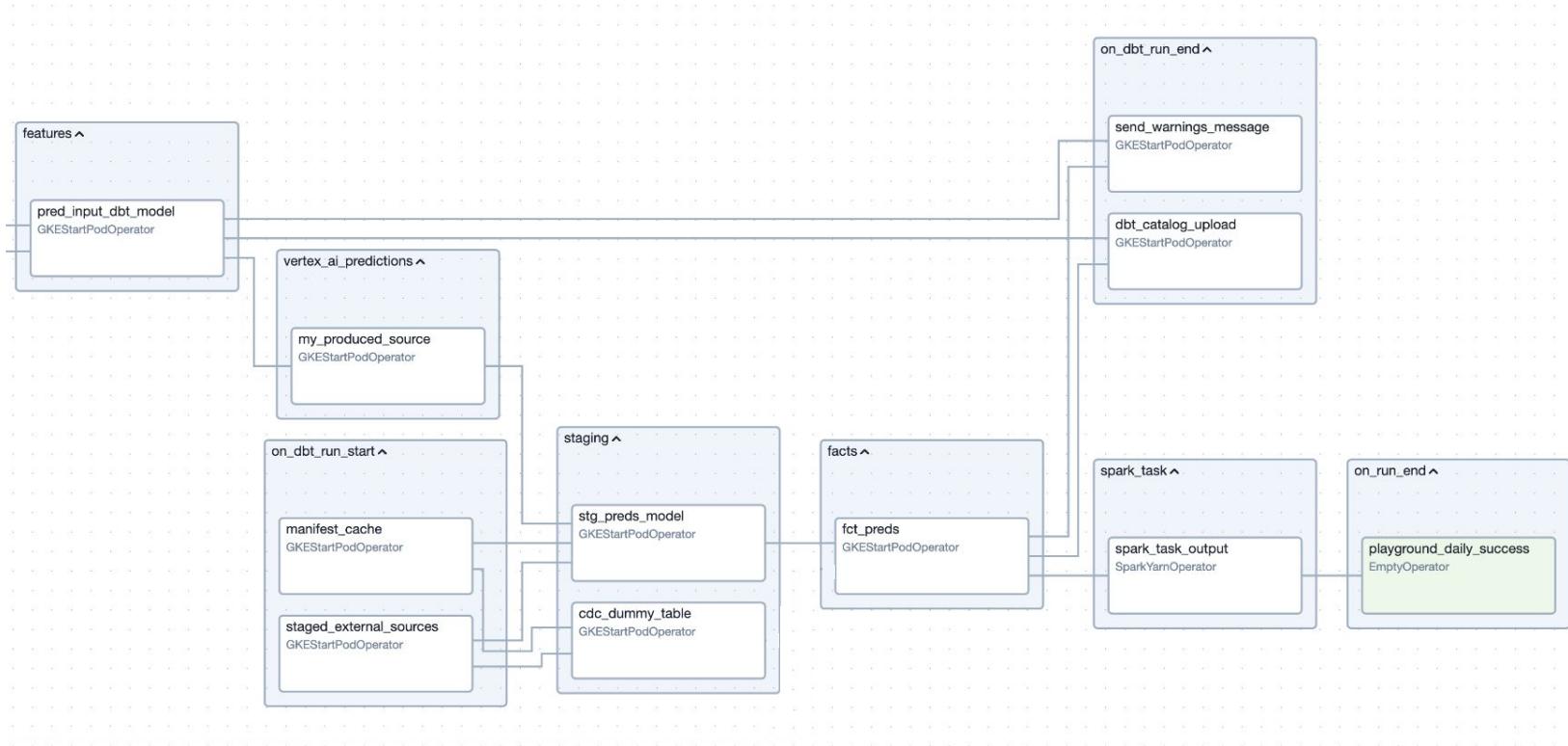






Generating DAGs

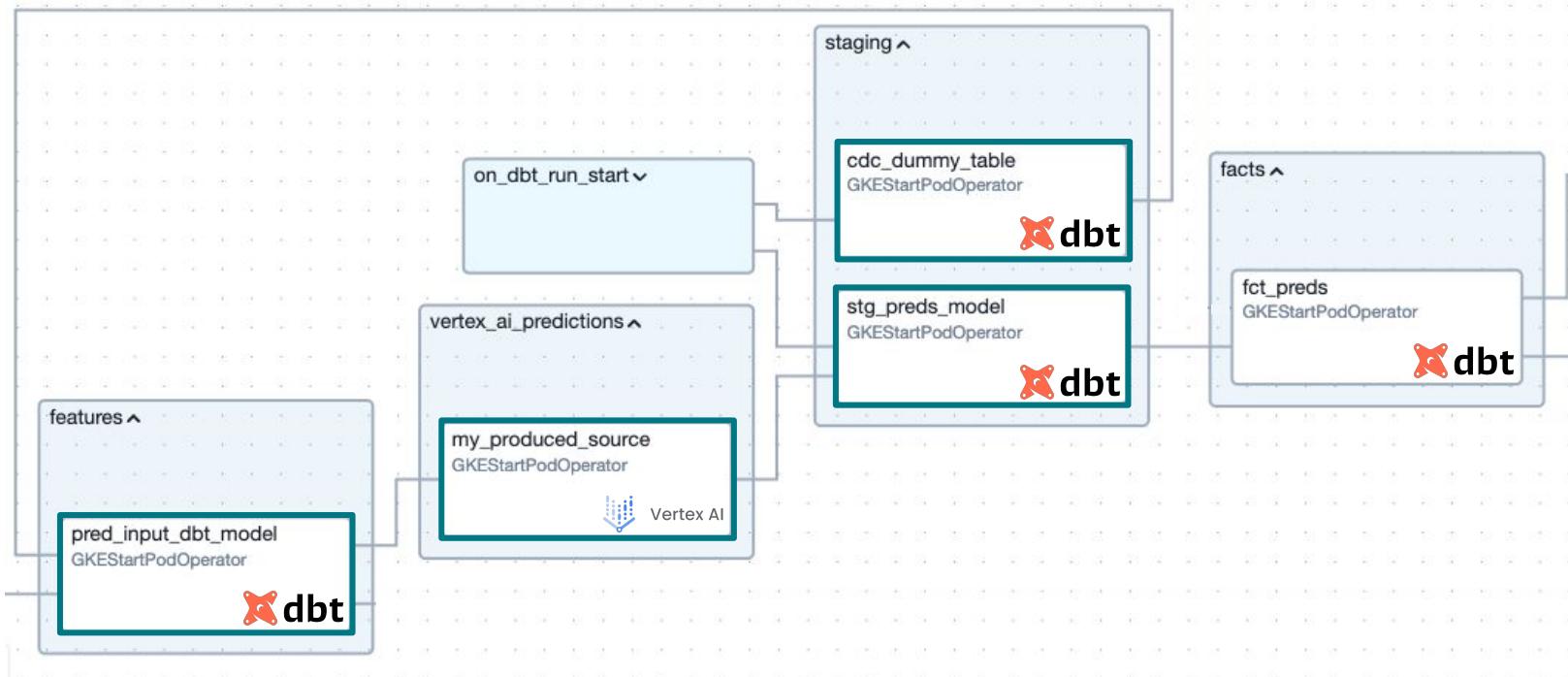




DAG generator design: User interface

```
1 version: 0.17
2
3 tags:
4   - tag_1
5   - long_tag
6
7 alerting:
8   channel: "#some-alerts-channel"
9   tag:
10    - "@amazing.person" # a person
11    - "!subteam^S05VAFP4A6N" # a Team
12
13 dbt_settings:
14   enabled: true
15   on_run_start:
16     - stage_external_sources
17   on_run_end:
18     - send_warnings_message
```

```
19 tasks:
20   - kind: universal
21     group: some_task_group_name
22     conditions:
23       - some_source
24     assets:
25       - produced_output_table
26     timeout:
27       minutes: 30
28     image: image_name:tag
29     command: poetry run python -m my_module.main
30
31 exposed_assets:
32   - regex: "mrt_daily_exposed"
33   slo:
34     daily:
35       hours: 4
```



Connecting Tasks

dbt_models.yml

```
1 models:  
2   - name: input_dbt_model  
3     config:  
4       meta:  
5         scheduler:  
6           pipelines:  
7             - hourly
```

scheduler.yaml

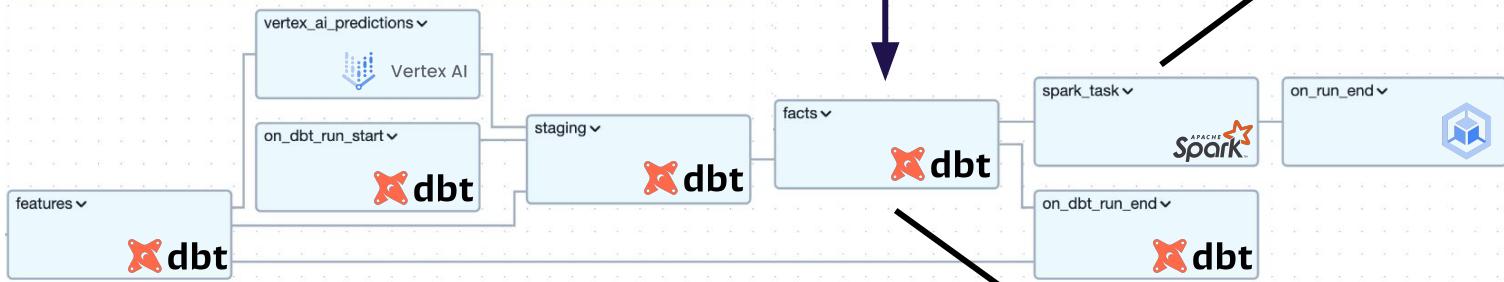
```
1 tasks:  
2   - kind: vertex-ai  
3     group: predictions  
4     conditions:  
5       - input_dbt_model  
6     assets:  
7       - my_produced_source
```

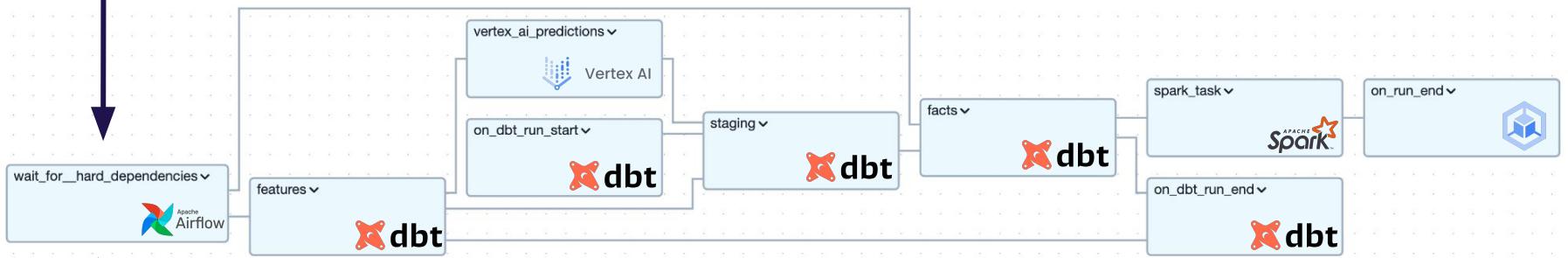
predictions_sources.yml

```
1 sources:  
2   - name: predictions  
3     tables:  
4       - name: my_produced_source  
5     columns:  
6       ...
```

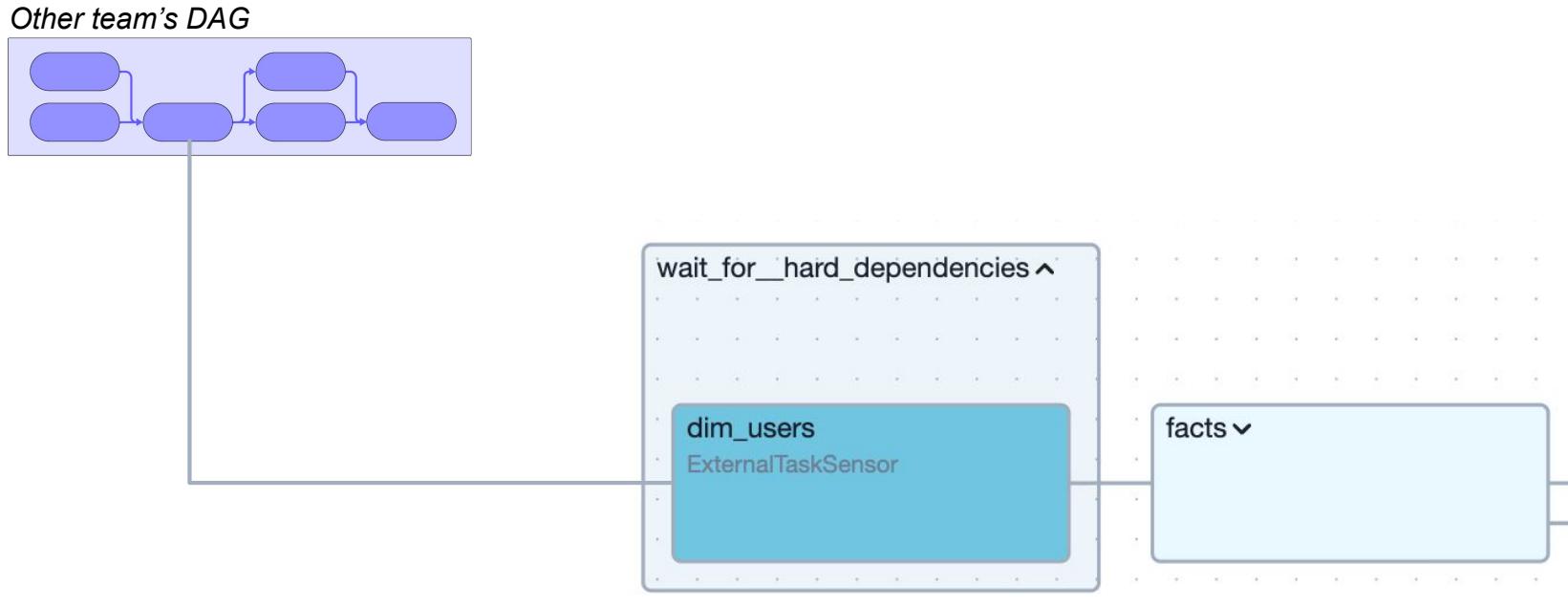
some_model.sql

```
1 ...  
2 select customer_id, predicted_value  
3 from {{ source("predictions", "my_produced_source") }}  
4 ...
```





Using other team's assets



Using other team's assets

scheduler.yaml

```
1 exposed_assets:  
2   # All dimensions  
3   - regex: "dim_.*"  
4     slo:  
5       daily:  
6         hours: 6
```

Asset registry object

```
1 {  
2   "some_team": {  
3     "produced": {  
4       "dim_customers": {  
5         "dag_prefix": "some_team",  
6         "task_group": "dimensions",  
7         "task_id": "dim_customers",  
8         "pipeline": "daily",  
9         "slo_seconds": 21600  
10      }  
11    }  
12  }
```



CI/CD validation

```
1 sources:  
2   - name: global_assets  
3     tables:  
4       - name: dim_customers
```

```
1 ...  
2 select customer_id, country_id  
3 from {{ source("predictions", "dim_customers") }}  
4 ...
```



Developer Flow

playground/models/global_assets/global_sources.yml

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

+4 Viewed

playground/models/marts/my_new_model.sql

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

+3 Viewed



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

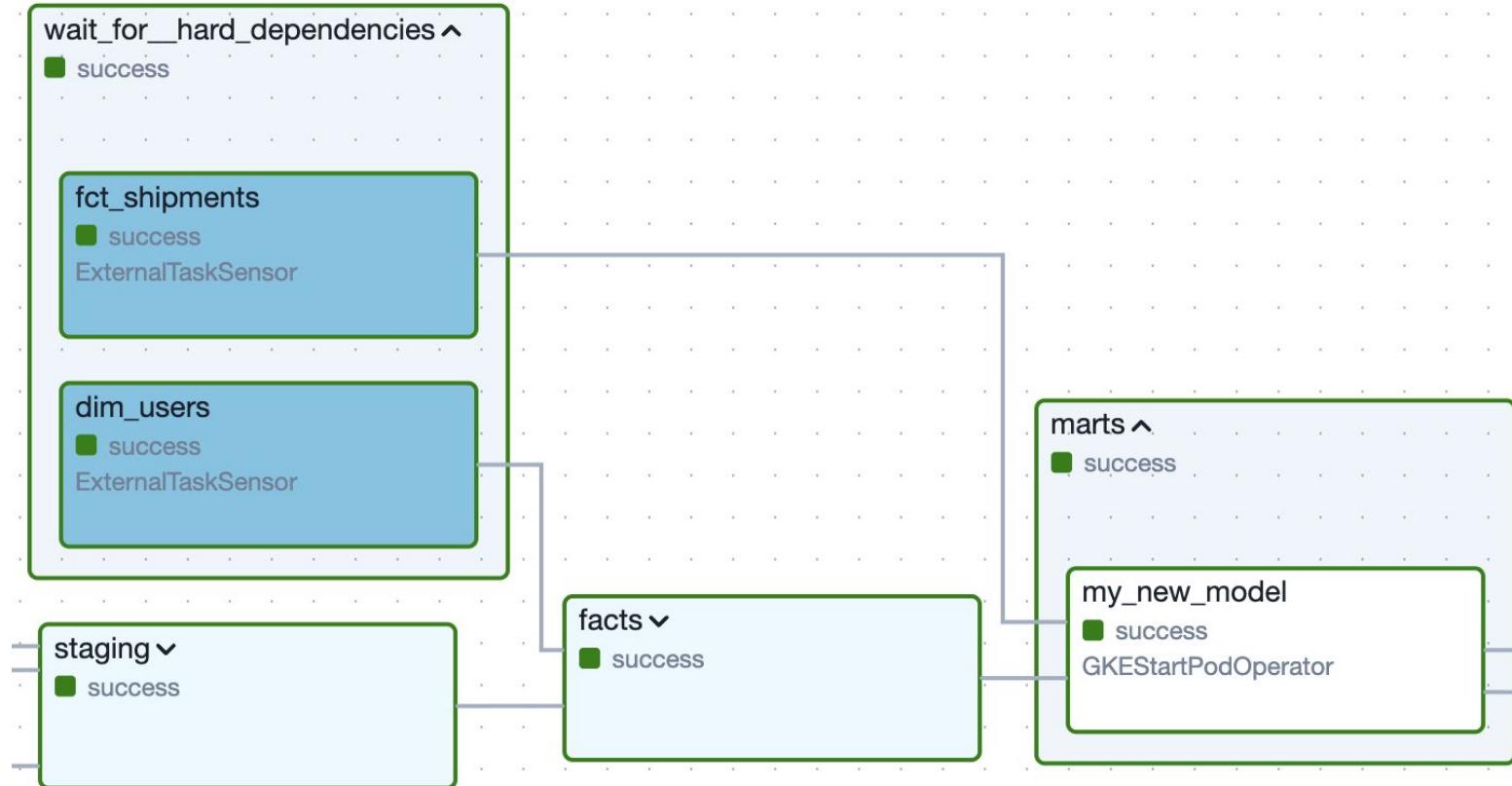
⚠️ Breaking changes found in project registry

Asset test_users

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

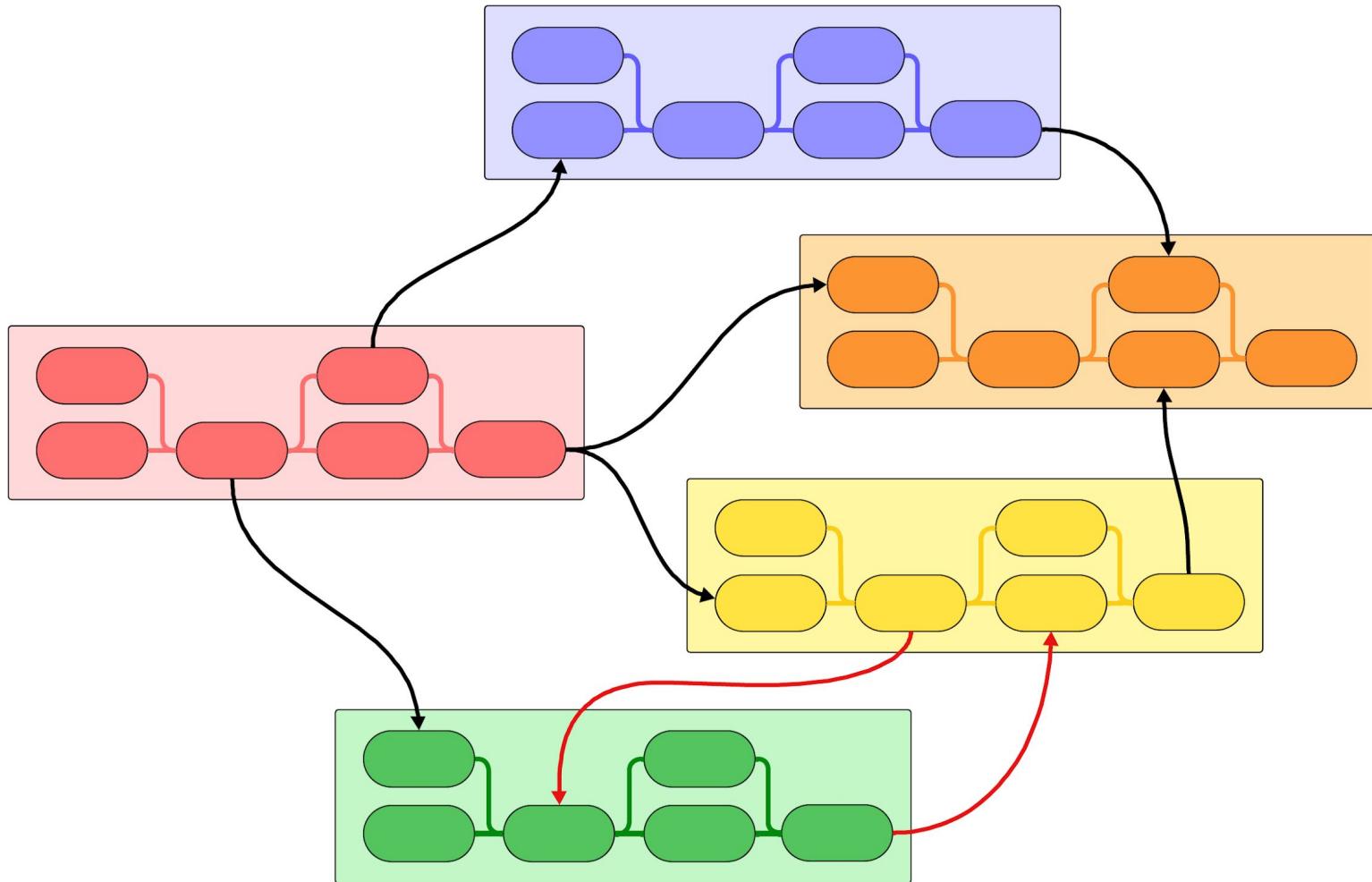
Consumers:

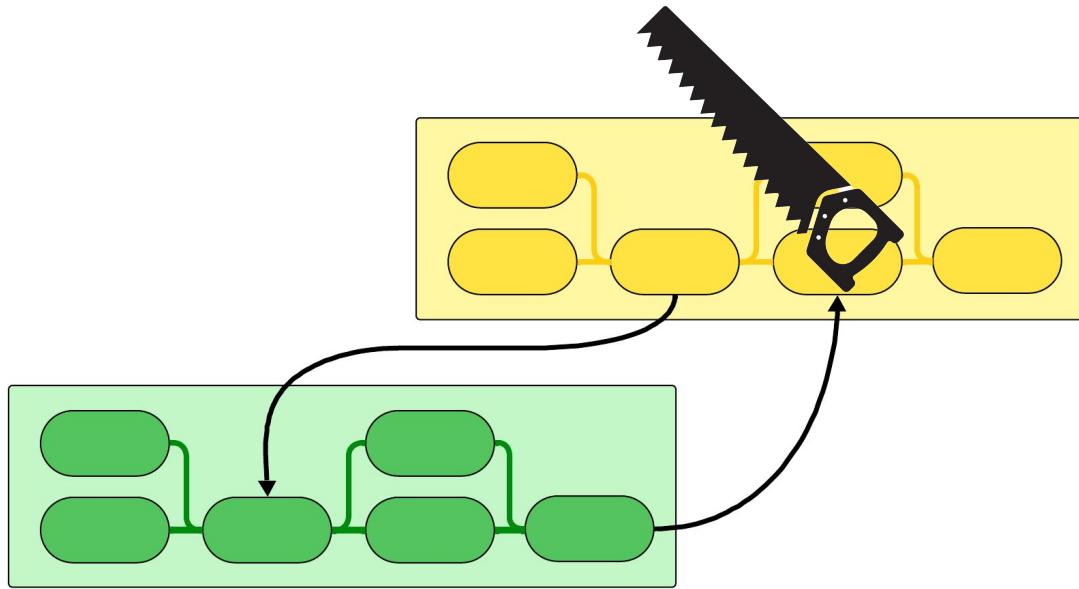
- *some_team:#some-team-contact-channel*

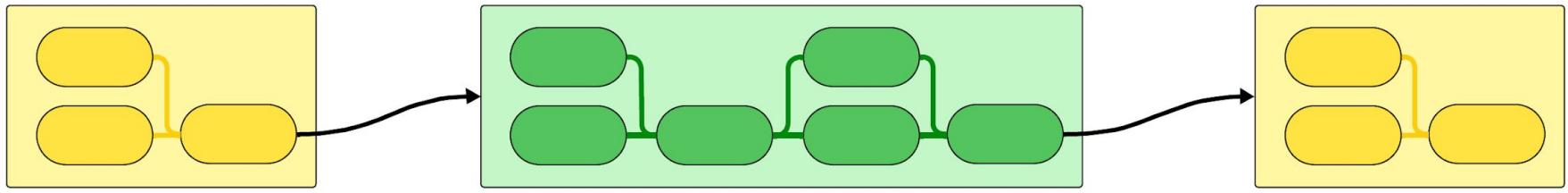


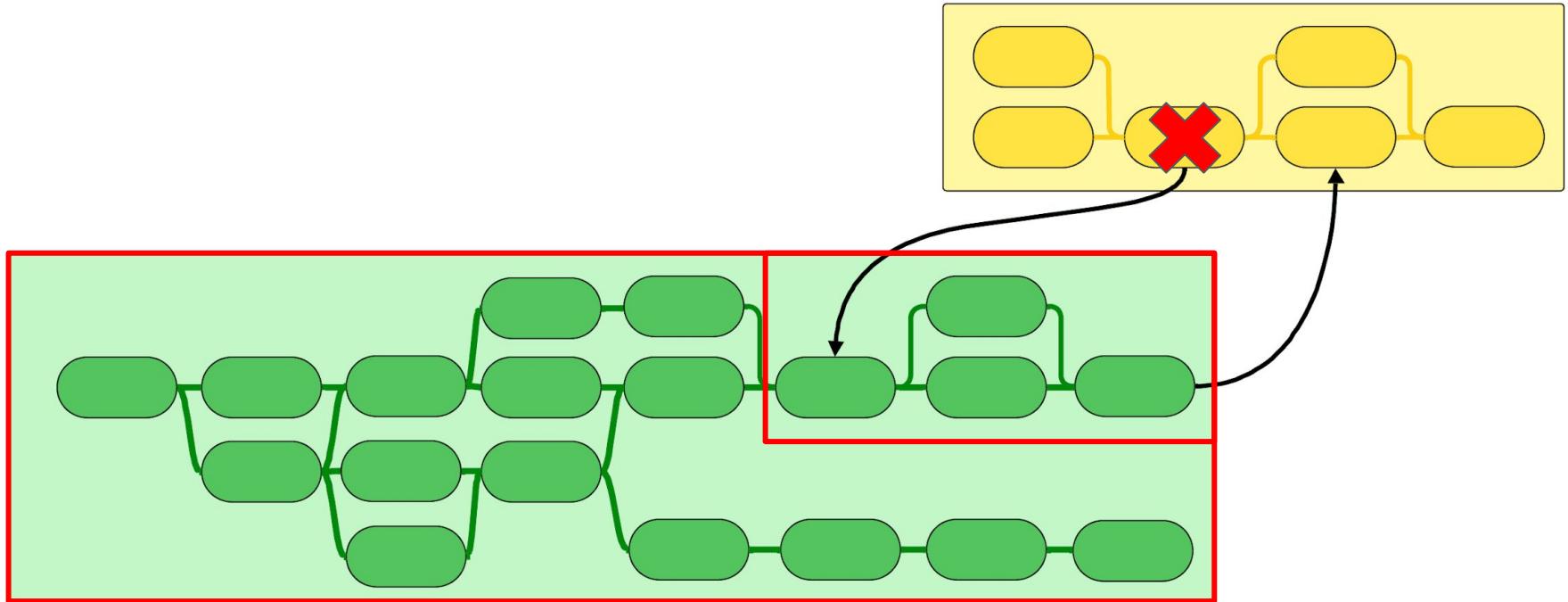
Sensor Setup











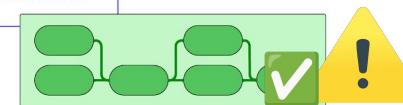
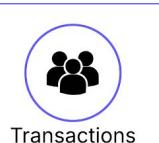


supply_daily task failure 🚧

Date interval start: 2025-09-09 00:10:00+00:00

[marts.mrt_my_model - view logs](#)

@some.owner.team, @some.owner.person



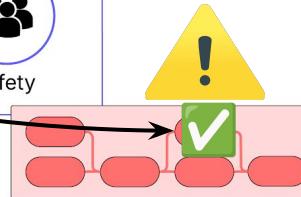
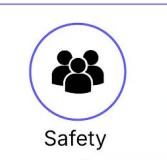
transactions_daily dependency failure 🚧

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



safety_daily dependency failure 🚧

Date interval start: 2025-09-09 00:10:00+00:00

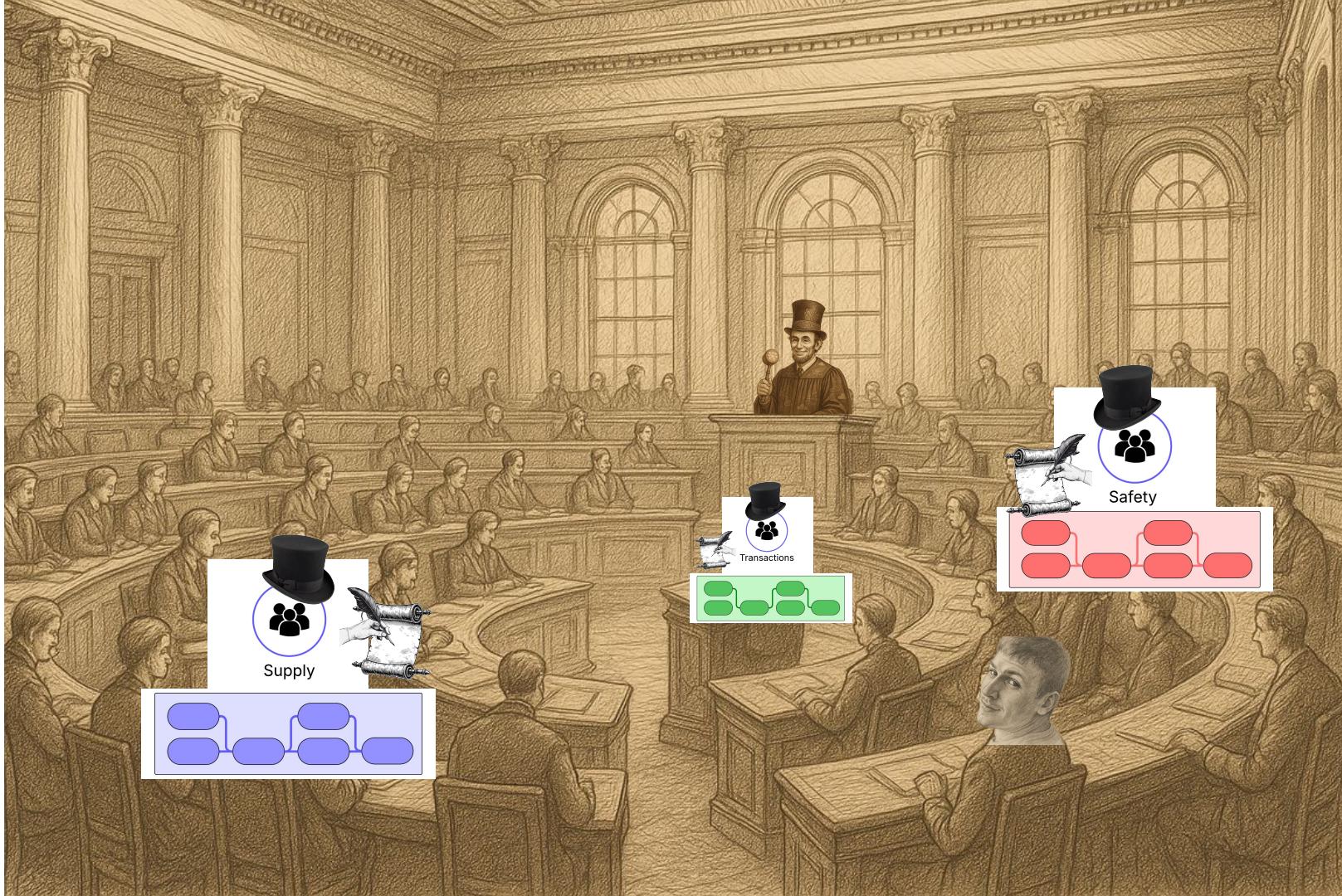
[wait_for__hard_dependencies.mrt_my_model - view logs](#)

When the task finishes successfully on the producers' side, your DAG will resume automatically.

@some.owner.team, @some.owner.person

The New Normal







Thank you!
Questions?