

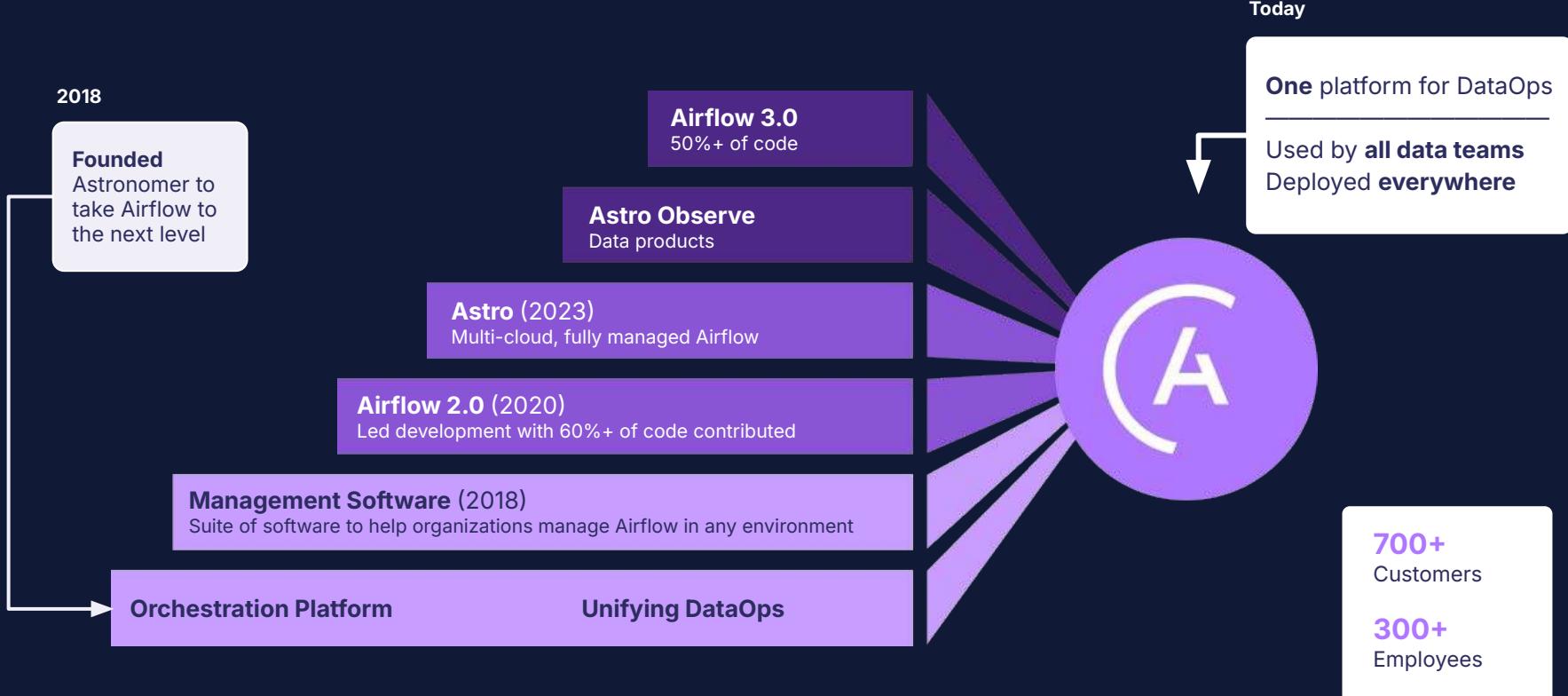


Orchestrating Data Quality - Quality Data Brought To You By Airflow

Maggie Stark

3.0

ASTRONOMER





- ⌚ Centralized Data Team
- ➡ Build critical operational and analytical pipelines with Airflow
- 👥 Product Influencers
- 📊 Providing a Data-First Perspective

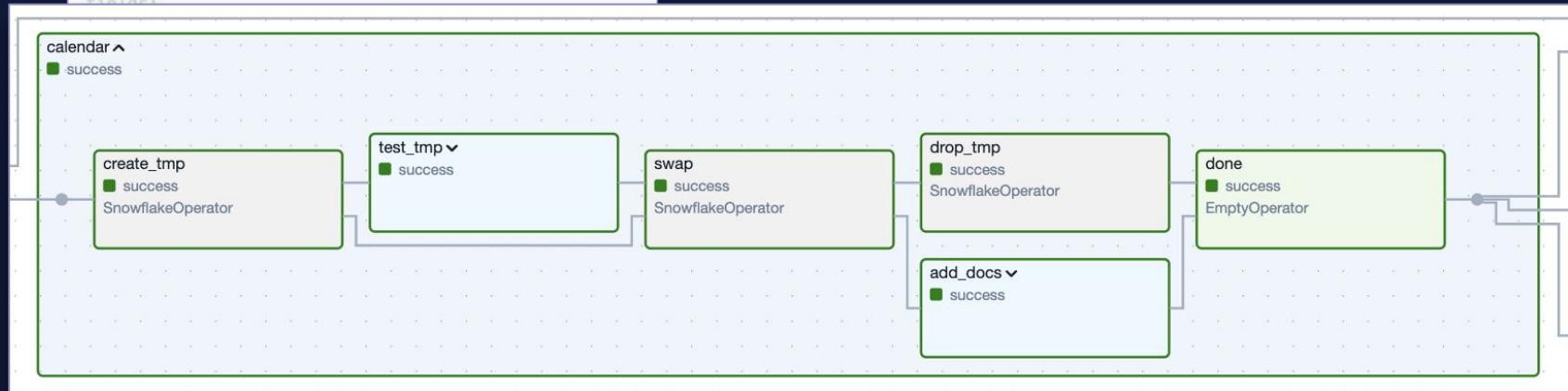
Configuration as Code

```
/*
operator: include.task_groups.transform.CreateTable
description: >-
    Creates standard calendar, where each row
    represents a date with common date operations.
fields:
    date: Date (iso format yyyy-mm-dd).
    is_weekend: Date is Sat. or Sun.
    is_weekday: Date is weekday (MTWTF).
schema: !switch_value
    sandbox: env_sandbox_schema
    default: commons
primary_key:
    - date
tests:
    check_null:
        - is_weekend
*/
SELECT
    gd.date::DATE AS date,
    IFF(gd.dow IN (0, 6), TRUE, FALSE) AS is_weekend,
    (NOT is_weekend) AS is_weekday
FROM generated_dates AS gd
ORDER BY date ASC
```

calendar.sql

Configuration as Code

```
/*
operator: include.task_groups.transform.CreateTable
description: >-
    Creates standard calendar, where each row
    represents a date with common date operations.
fields:
```

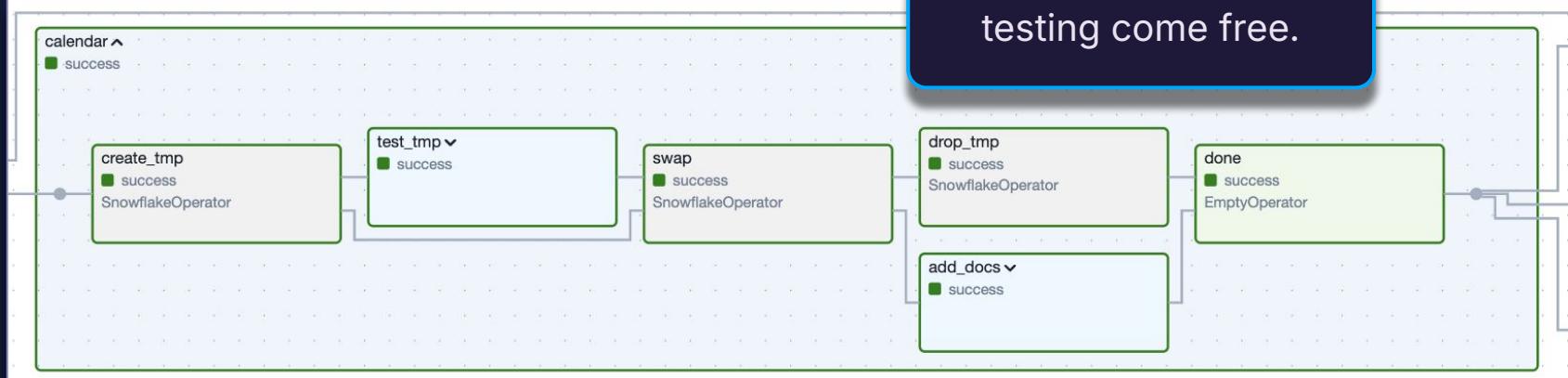


```
gd.date::DATE AS date,
IFF(gd.dow IN (0, 6), TRUE, FALSE) AS is_weekend,
(NOT is_weekend) AS is_weekday
FROM generated_dates AS gd
ORDER BY date ASC
```

calendar.sql

Configuration as Code

```
/*
operator: include.task_groups.transform.CreateTable
description: >-
    Creates standard calendar, where each row
    represents a date with common date operations.
fields:
```

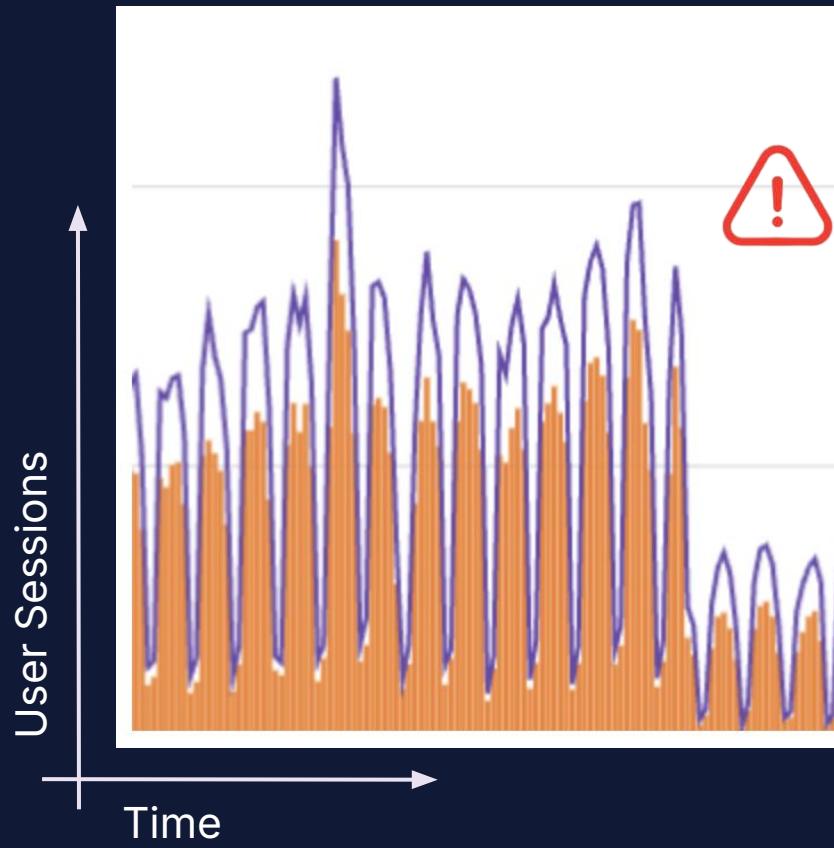


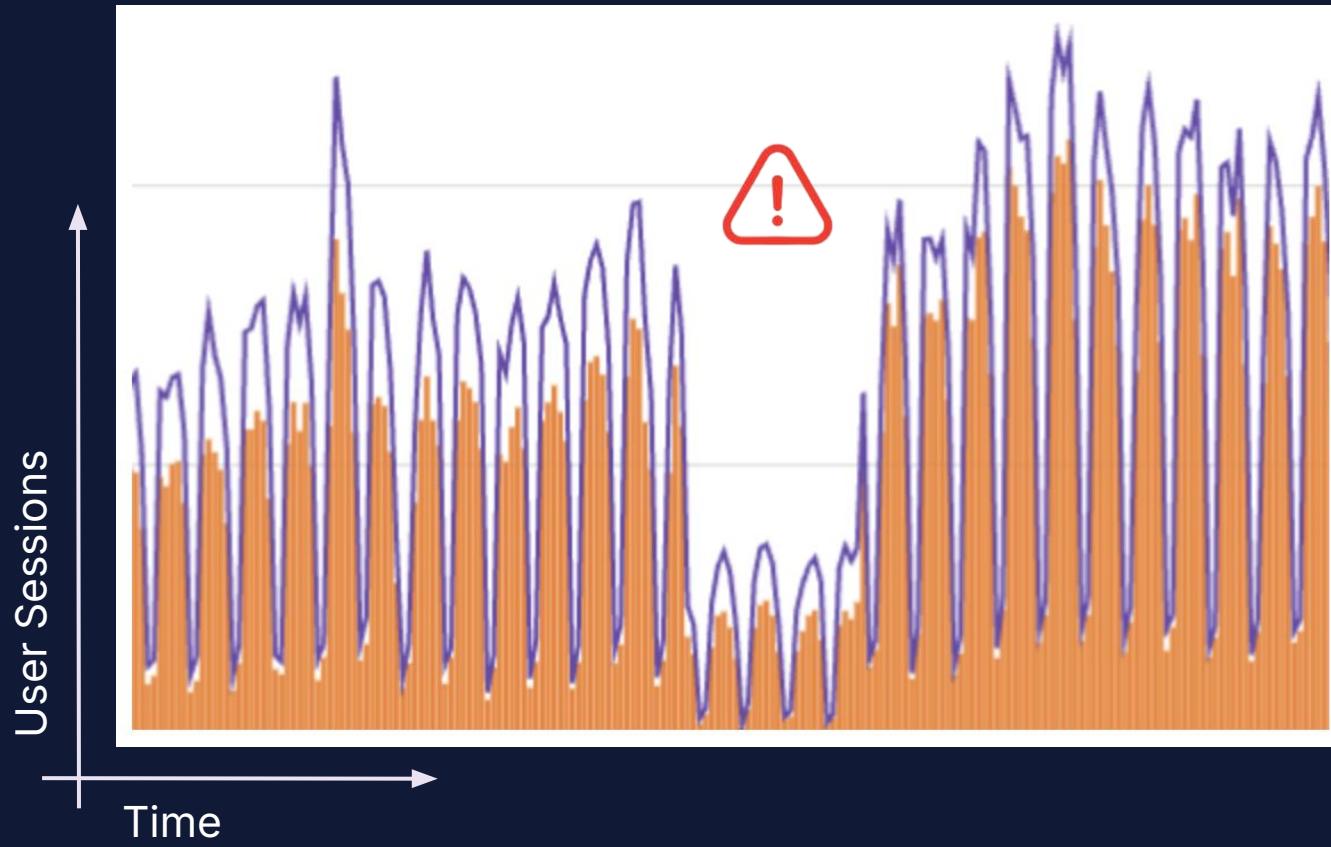
```
gd.date::DATE AS date,
IFF(gd.dow IN (0, 6), TRUE, FALSE) AS is_weekend,
(NOT is_weekend) AS is_weekday
FROM generated_dates AS gd
ORDER BY date ASC
```

calendar.sql

1 file = n tasks!

Create a table; get documentation and testing come free.



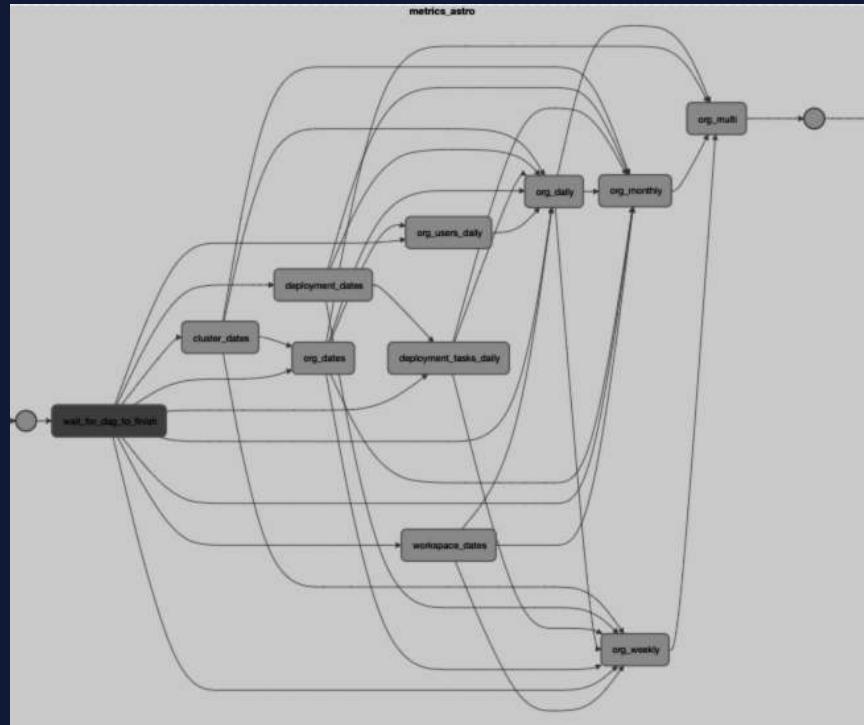


- Standalone Data-Quality DAG

- Hard to Maintain

- Duplicative

- Poor Visibility and Adoption



Configuration as Code

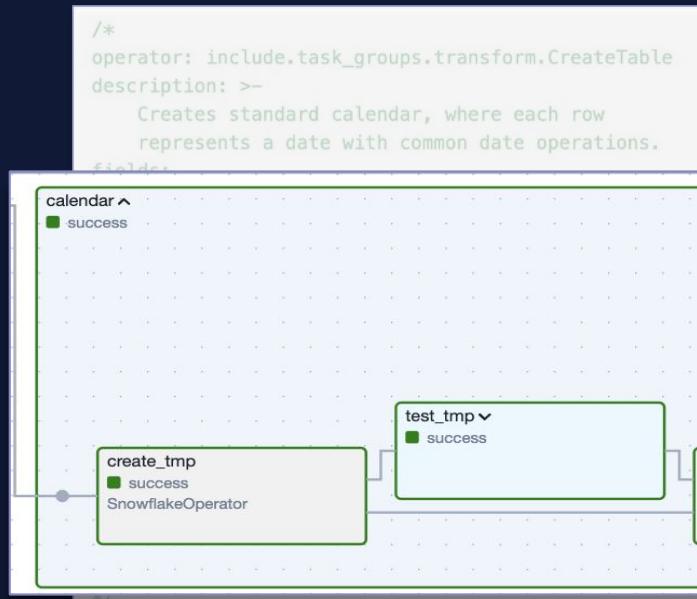
```
/*
operator: include.task_groups.transform.CreateTable
description: >-
    Creates standard calendar, where each row
    represents a date with common date operations.
fields:
    date: Date (iso format yyyy-mm-dd).
    is_weekend: Date is Sat. or Sun.
    is_weekday: Date is weekday (MTWTF).
schema: !switch_value
    sandbox: env_sandbox_schema
    default: commons
primary_key:
    - date
tests:
    check_null:
        - is_weekend
validations:
    check_condition:
        - is_weekend != is_weekday
*/

```

```
SELECT
    gd.date::DATE AS date,
    IFF(gd.dow IN (0, 6), TRUE, FALSE) AS is_weekend,
    (NOT is_weekend) AS is_weekday
FROM generated_dates AS gd
ORDER BY date ASC
```

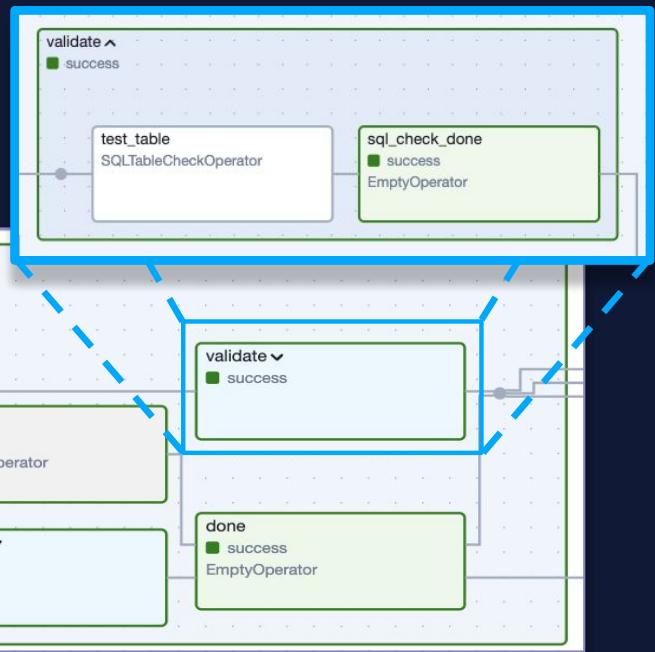
calendar.sql

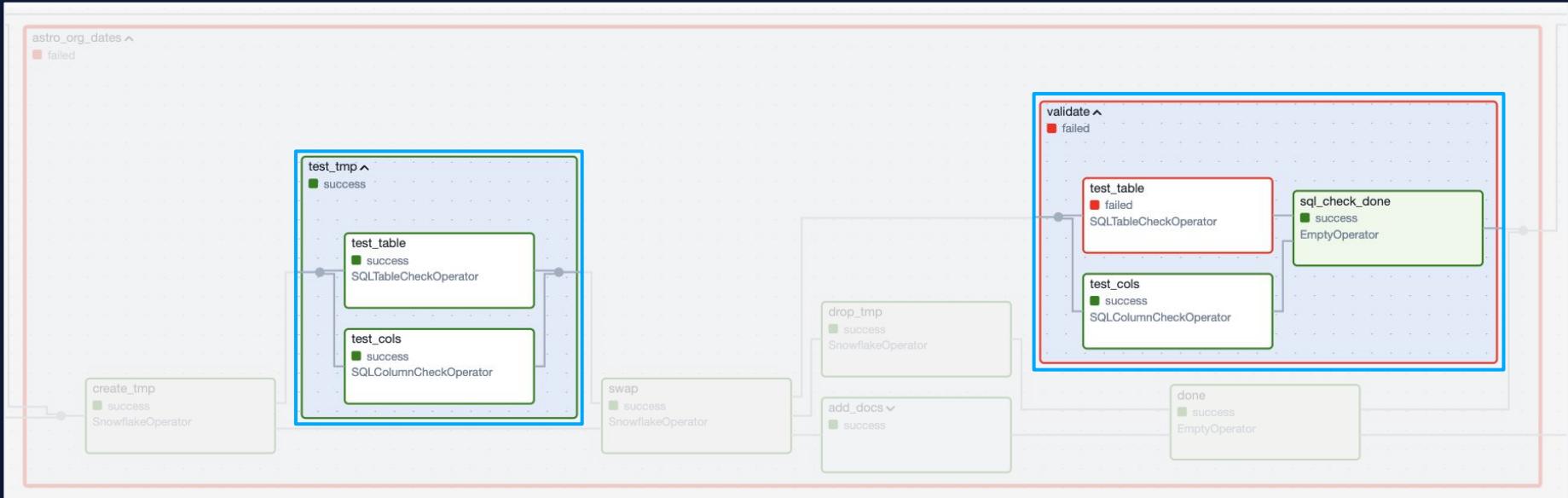
Configuration as Code



```
SELECT
    gd.date::DATE AS date,
    IFF(gd.dow IN (0, 6), TRUE, FALSE) AS is_weekend,
    (NOT is_weekend) AS is_weekday
FROM generated_dates AS gd
ORDER BY date ASC
```

calendar.sql





Data Contracts

Expected attributes exist, NULLs, field values, conditionals based on understanding of data source.

Data Integrity

PKs, FKs, NULLs, unique field composites, field values.

Data Accuracy

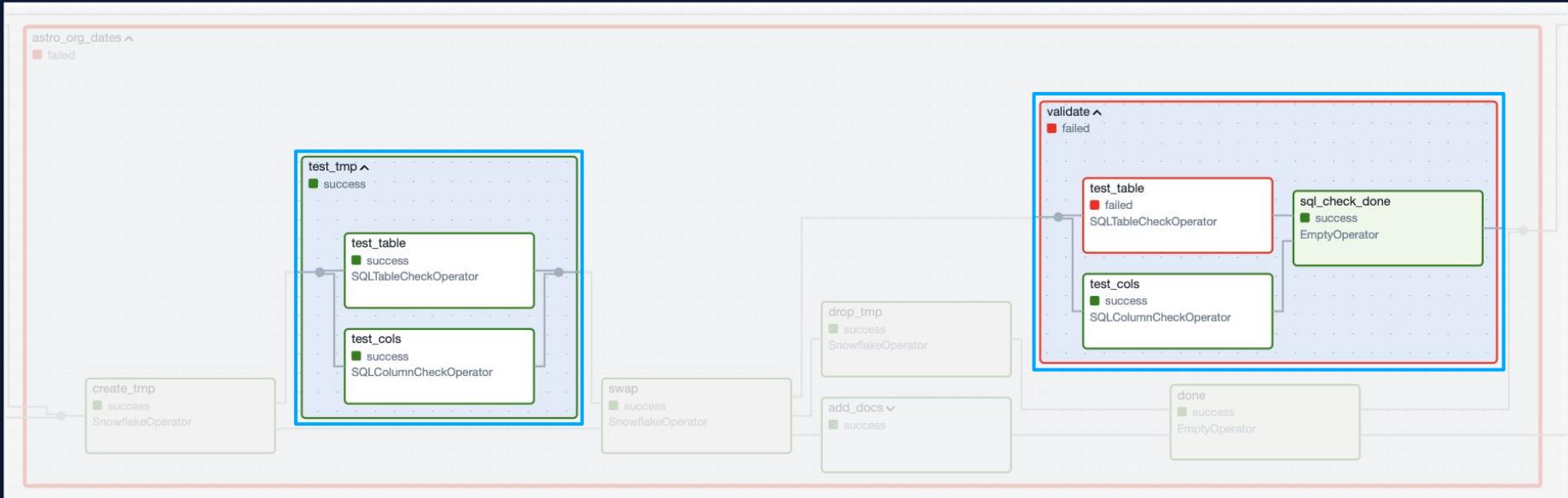
Metrics align between related grains of aggregations, business-process conditionals which have external impact.

Non-Blocking Validations

Business-process conditionals with no external impact, field values, parity between dev./prod.

Where are data quality checks performed?





```
/*
operator: include.task_groups.transform.CreateTable
description: >-
    Creates standard calendar table
    represents a date with week day
fields:
    date: Date (iso format)
    is_weekend: Date is Saturday or Sunday
    is_weekday: Date is Monday through Friday
schema: !switch_value
    sandbox: env_sandbox
    default: commons
primary_key:
    - date
tests:
    check_null:
        - is_weekend
        - is_weekday
validations:
    check_condition:
        - is_weekend != is_weekday
*/

```

validations:

check_condition:

external:

- date = eom_date

internal:

- is_weekend OR is_weekday
- COUNT(*) = 365

```
SELECT
    gd.date::DATE AS date,
    IFF(gd.dow IN (0, 6), TRUE, FALSE) AS is_weekend,
    (NOT is_weekend) AS is_weekday
FROM generated_dates AS gd
ORDER BY date ASC
```

calendar.sql

Logging

	DAG_ID	TASK_ID	LOGICAL_TS
1	model_crm	sf_accounts.validate.test_table	2025-09-18 06:29:44

```
[{"RESULTS": [{"failed_row_count": "50", "is_success": "False", "owner": "external", "sql": "zd_org_id IS NOT NULL OR is_deleted OR NOT is_current_cust", "test_name": "check_condition_1"}, {"failed_row_count": "10", "is_success": "False", "owner": "external", "sql": "NOT is_current_cust OR support_level IS NOT NULL OR product IS NULL", "test_name": "check_condition_2"}]}]
```

	DAG_ID	TASK_ID	LOGICAL_TS
1	test_mart_cust	test_current_astro_custs	2025-09-17 17:46:10

```
[{"RESULTS": {"owner": "internal", "test_severity": "validation", "failed_row_count": 2, "is_success": false, "failed_rows": "D566485B7B22414343545F4944223A22303031347830010124627653357A414145222C091F4C4E414D45223A26E647265645820496E632E111CC454414753223A2250726F6A6563746564207573616765206F76657220313235F6E73756D7074696F6E2067726F777468091C4035302520284C61737420372064617973298A2A000433300D2B5772075736572732061646465643B41637469766500132469657211174C55524C223A2268747470733A2F2F617374726F62E636F6D2F1514282F722F4163636F756E742F463601102F733044223A395A1A00185F4C4147223A32361E000450430510"}, {"HUMAN_READABLE": [{"ACCT_ID": "0014x00000bvS5zAAE"}]}]
```



Recapping

-  Use Airflow-native operators to perform DQ checks.
-  Write-audit-publish approach.
-  Checks are mainly configured in task definition
 -  New/different standalone DQ pipelines
-  Surface some validation results in dashboards to take action.

The 2025 Apache Airflow® Survey is here!

Fill it out to for a free Airflow 3
Fundamentals or DAG Authoring in
Airflow 3 certification code



Questions?