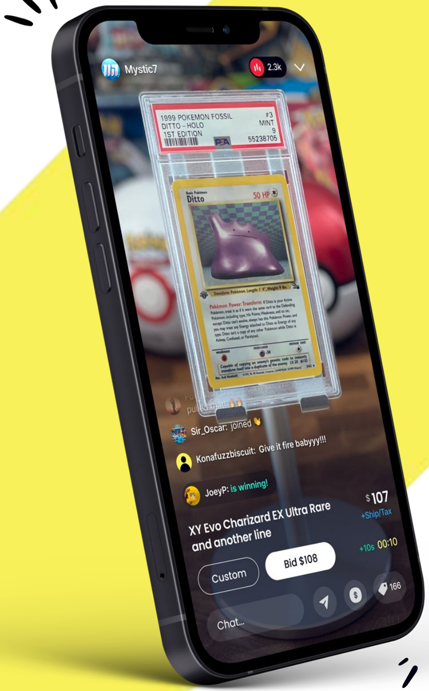
Extreme Self-Service:
Turning Data Consumers into Data Constructors

Alice Leach whatnot



Sir_Oscar: joined 👋

Konafuzzbiscuit: Give it fire babyyy!!!

JoeyP: is winning!

XY Evo Charizard EX Ultra Rare and another line

Whatnot

Buy, Sell & Go Live

Live shopping platform and marketplace

- Founded in Dec 2019
- Fastest growing marketplace in the US 2 years in a row (Marketplace 100)
- 80+ product categories

Self-Service Data Platform

Regular

- Access data
- Run queries
- Build dashboards
- Generate reports

Extreme

- Build new datasets
- Productionize reports
- Generate new data
- Push data to desired tools

Build community & SQL literacy. Establish & share best practices. Build robust guard rails.

Consumers & Constructors

Business Stakeholders

- Category success
- Buyer/ seller growth

Analytics/ Machine Learning

- Feed ranking
- Customer engagement

Software Engineers

- Chat tagging adoption
- Seller dashboard

Data Insight Developers

- Data assets
- Snowflake, Sigma, Retool
- SQL

Data Product Developers

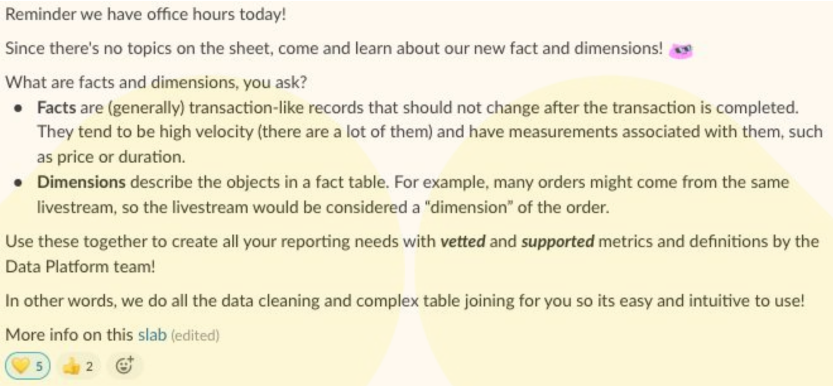
- Data products
- dbt, event-schema-registry
- Advanced SQL

Data Pipeline Developers

- Data processes
- dagster
- Python

Data Platform: Define and minimize interactive surfaces. Unblock users on those surfaces.

Community & SQL Literacy



Reminder we have office hours today!

Since there's no topics on the sheet, come and learn about our new fact and dimensions!

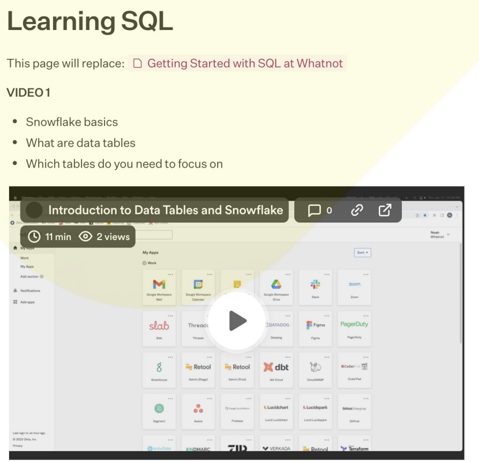
What are facts and dimensions, you ask?

- **Facts** are (generally) transaction-like records that should not change after the transaction is completed. They tend to be high velocity (there are a lot of them) and have measurements associated with them, such as price or duration.
- **Dimensions** describe the objects in a fact table. For example, many orders might come from the same livestream, so the livestream would be considered a "dimension" of the order.

Use these together to create all your reporting needs with ***vetted*** and ***supported*** metrics and definitions by the Data Platform team!

In other words, we do all the data cleaning and complex table joining for you so its easy and intuitive to use!

More info on this slab (edited)

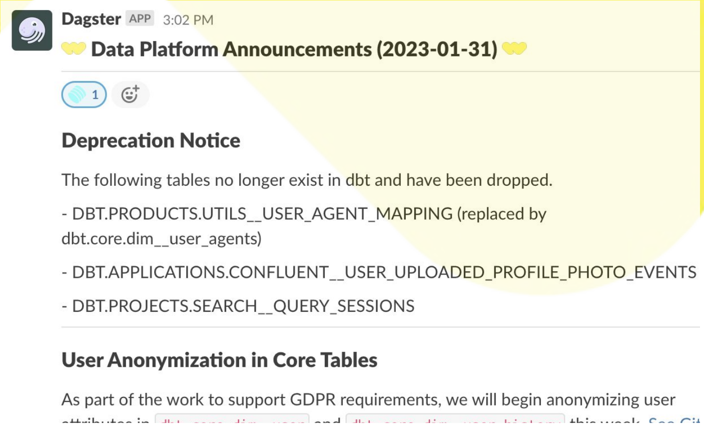


Learning SQL

This page will replace: Getting Started with SQL at Whatnot

VIDEO 1

- Snowflake basics
- What are data tables
- Which tables do you need to focus on



Dagster APP 3:02 PM

Data Platform Announcements (2023-01-31)

Deprecation Notice

The following tables no longer exist in dbt and have been dropped.

- DBT.PRODUCTS.UTILS__USER_AGENT_MAPPING (replaced by dbt.core.dim__user_agents)

- DBT.APPLICATIONS.CONFLUENT__USER_UPLOADED_PROFILE_PHOTO_EVENTS

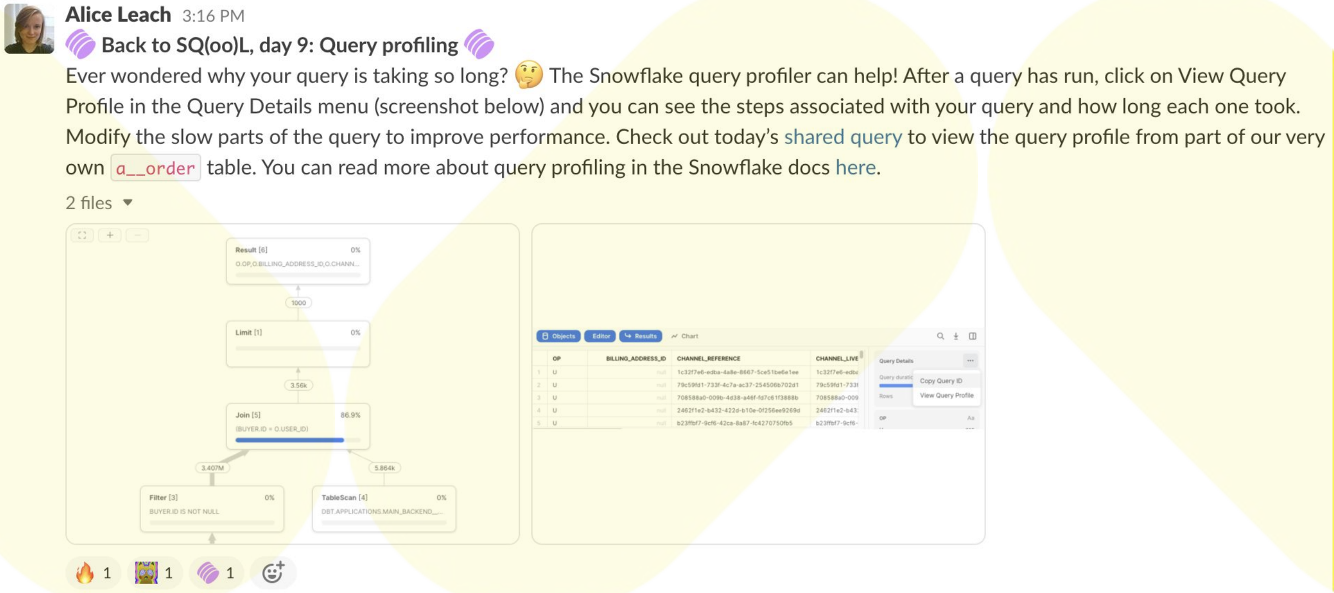
- DBT.PROJECTS.SEARCH__QUERY_SESSIONS

User Anonymization in Core Tables

As part of the work to support GDPR requirements, we will begin anonymizing user attributes in `dbt.core.dim__user` and `dbt.core.dim__user_history` this week. See Git

- Executive support
- Educational content
- Office hours
- Bank of common SQL queries
- Weekly data announcements

Case Study: Back to SQ(oo)L



Alice Leach 3:16 PM

Back to SQ(oo)L, day 9: Query profiling

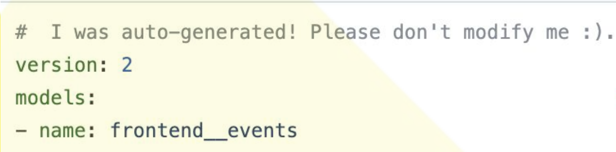
Ever wondered why your query is taking so long? 🤔 The Snowflake query profiler can help! After a query has run, click on View Query Profile in the Query Details menu (screenshot below) and you can see the steps associated with your query and how long each one took. Modify the slow parts of the query to improve performance. Check out today's shared query to view the query profile from part of our very own `a__order` table. You can read more about query profiling in the Snowflake docs here.

2 files ▾

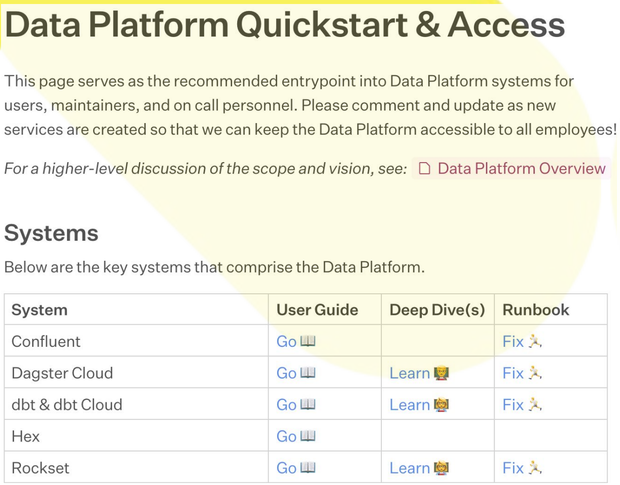
	OP	BILLING_ADDRESS_ID	CHANNEL_REFERENCE	CHANNEL_LIVE
1	U	null	1c32f7e6-edba-4a8e-8667-5ce51be6e1ee	1c32f7e6-edba
2	U	null	79c59fd1-733f-4c7a-ac37-254506b702d1	79c59fd1-733f
3	U	null	708588a0-009b-4d38-a46f-fd7c61f3888b	708588a0-009
4	U	null	2462f1e2-b432-422d-b10e-0f256ee9269d	2462f1e2-b43
5	U	null	b23ffbf7-9cf6-42ca-8a87-fc4270750fb5	b23ffbf7-9cf6-

Month long initiative. SQL tip a day shared by members of data, analytics and ML teams.

Best Practices



```
#  I was auto-generated! Please don't modify me :).
version: 2
models:
- name: frontend__events
```



Data Platform Quickstart & Access

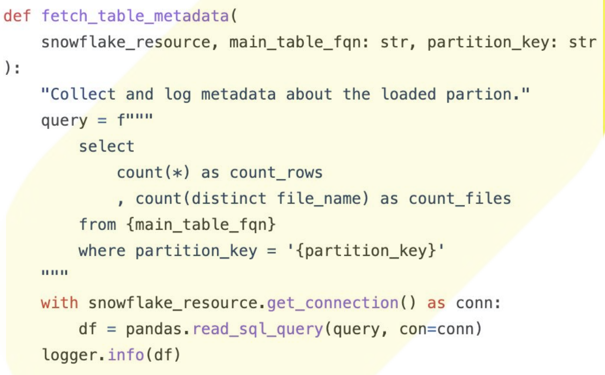
This page serves as the recommended entrypoint into Data Platform systems for users, maintainers, and on call personnel. Please comment and update as new services are created so that we can keep the Data Platform accessible to all employees!

For a higher-level discussion of the scope and vision, see: 🗋 Data Platform Overview

Systems

Below are the key systems that comprise the Data Platform.

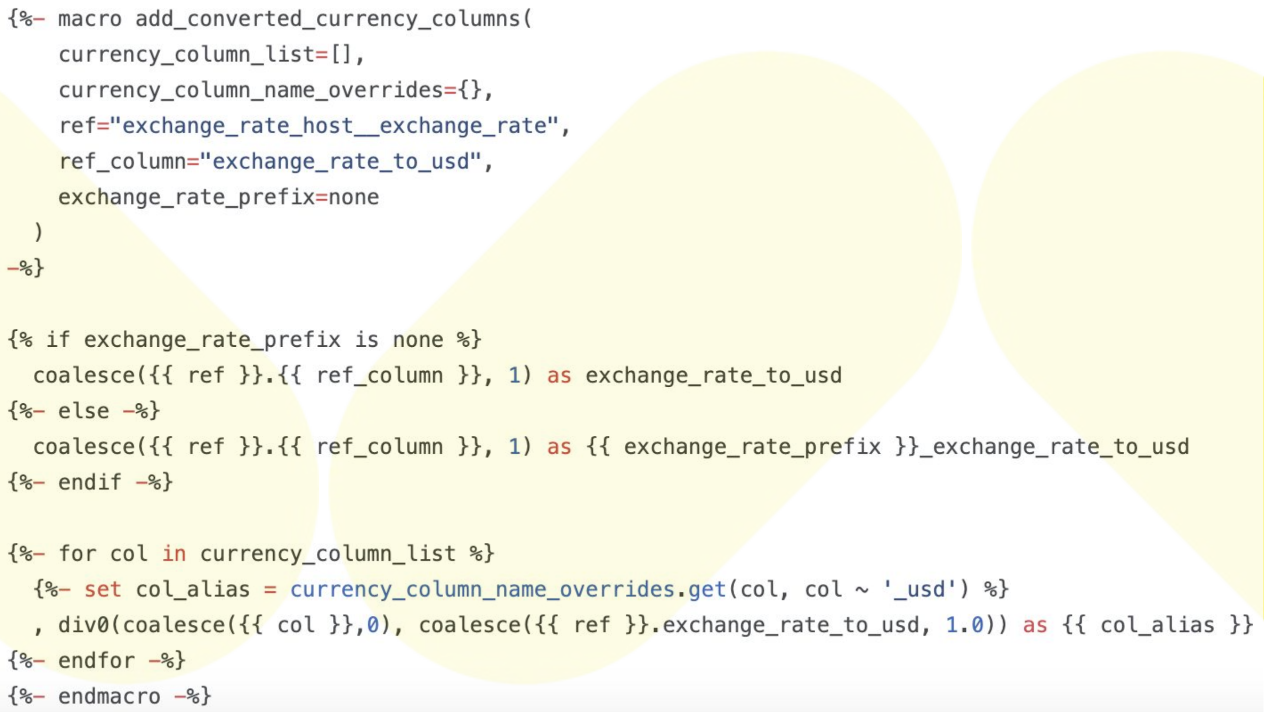
System	User Guide	Deep Dive(s)	Runbook
Confluent	Go 📖		Fix 🏃
Dagster Cloud	Go 📖	Learn 🧑‍🏫	Fix 🏃
dbt & dbt Cloud	Go 📖	Learn 🧑‍🏫	Fix 🏃
Hex	Go 📖		
Rockset	Go 📖	Learn 🧑‍🏫	Fix 🏃



```python
def fetch_table_metadata(
    snowflake_resource, main_table_fqn: str, partition_key: str
):
    "Collect and log metadata about the loaded partion."
    query = f"""
        select
            count(*) as count_rows
            , count(distinct file_name) as count_files
        from {main_table_fqn}
        where partition_key = '{partition_key}'
    """
    with snowflake_resource.get_connection() as conn:
        df = pandas.read_sql_query(query, con=conn)
    logger.info(df)
```

- Lead by example
  - SLAs
  - Core data marts
  - Utilities/ Macros
- Streamline schema changes
  - Protocol buffers
  - Codegen
- Templates/ Documentation
  - Example PRs
  - Wikipedia/ How-Tos
- Observability and testing
  - Slack-based alerting
  - Opt in defaults
- Own as little as possible

Case Study: Currency Conversion



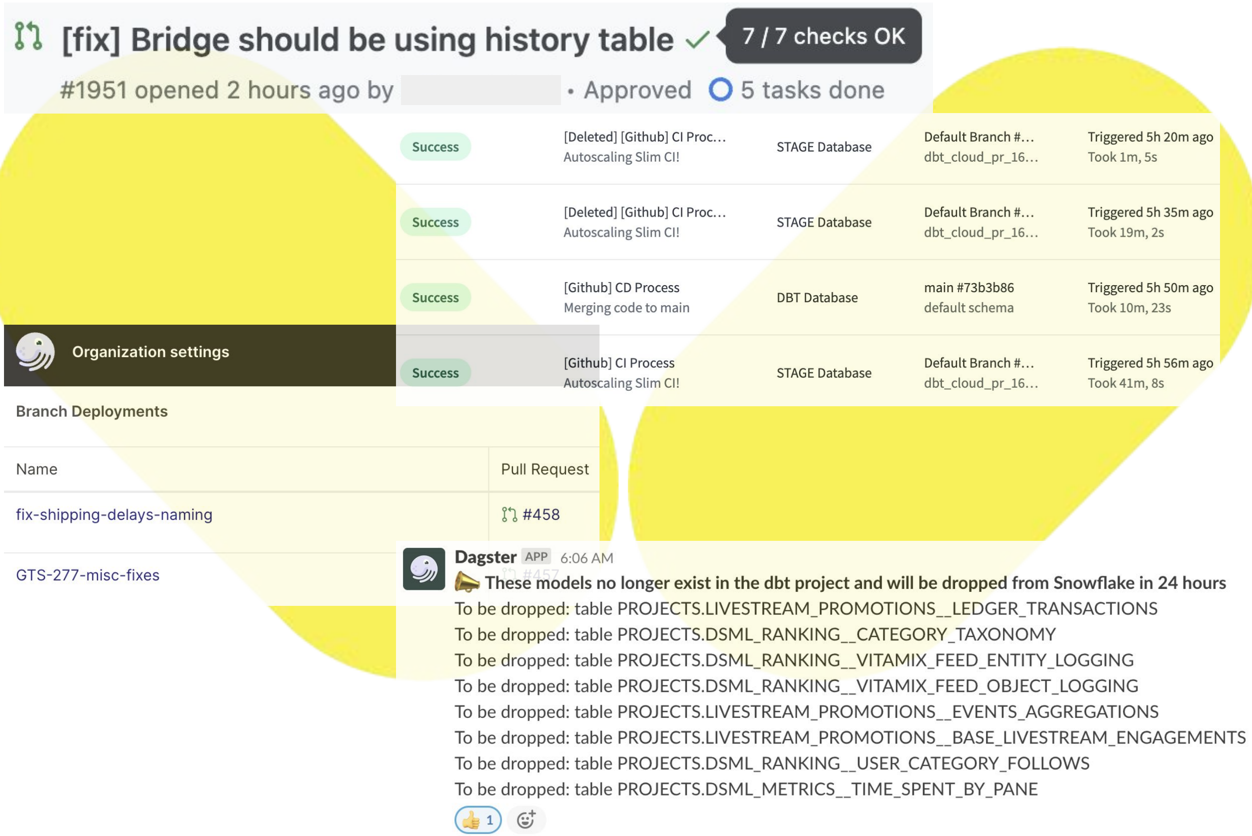
```
{%- macro add_converted_currency_columns(
    currency_column_list=[],
    currency_column_name_overrides={},
    ref="exchange_rate_host__exchange_rate",
    ref_column="exchange_rate_to_usd",
    exchange_rate_prefix=none
  )
-%}

{% if exchange_rate_prefix is none %}
  coalesce({{ ref }}.{{ ref_column }}, 1) as exchange_rate_to_usd
{%- else -%}
  coalesce({{ ref }}.{{ ref_column }}, 1) as {{ exchange_rate_prefix }}_exchange_rate_to_usd
{%- endif -%}

{%- for col in currency_column_list %}
  {%- set col_alias = currency_column_name_overrides.get(col, col ~ '_usd') %}
  , div0(coalesce({{ col }},0), coalesce({{ ref }}.exchange_rate_to_usd, 1.0)) as {{ col_alias }}
{%- endfor -%}
{%- endmacro -%}
```

SQL macro to convert and name price fields from multi currency to USD.

Robust Guard Rails



[fix] Bridge should be using history table ✓ 7 / 7 checks OK

#1951 opened 2 hours ago by [redacted] • Approved ○ 5 tasks done

Success	[Deleted] [Github] CI Proc… Autoscaling Slim CI!	STAGE Database	Default Branch #… dbt_cloud_pr_16…	Triggered 5h 20m ago Took 1m, 5s
Success	[Deleted] [Github] CI Proc… Autoscaling Slim CI!	STAGE Database	Default Branch #… dbt_cloud_pr_16…	Triggered 5h 35m ago Took 19m, 2s
Success	[Github] CD Process Merging code to main	DBT Database	main #73b3b86 default schema	Triggered 5h 50m ago Took 10m, 23s
Success	[Github] CI Process Autoscaling Slim CI!	STAGE Database	Default Branch #… dbt_cloud_pr_16…	Triggered 5h 56m ago Took 41m, 8s

Branch Deployments

Name	Pull Request
fix-shipping-delays-naming	#458
GTS-277-misc-fixes	#457

Dagster APP 6:06 AM

📣 **These models no longer exist in the dbt project and will be dropped from Snowflake in 24 hours**

To be dropped: table PROJECTS.LIVESTREAM_PROMOTIONS__LEDGER_TRANSACTIONS

To be dropped: table PROJECTS.DSML_RANKING__CATEGORY_TAXONOMY

To be dropped: table PROJECTS.DSML_RANKING__VITAMIX_FEED_ENTITY_LOGGING

To be dropped: table PROJECTS.DSML_RANKING__VITAMIX_FEED_OBJECT_LOGGING

To be dropped: table PROJECTS.LIVESTREAM_PROMOTIONS__EVENTS_AGGREGATIONS

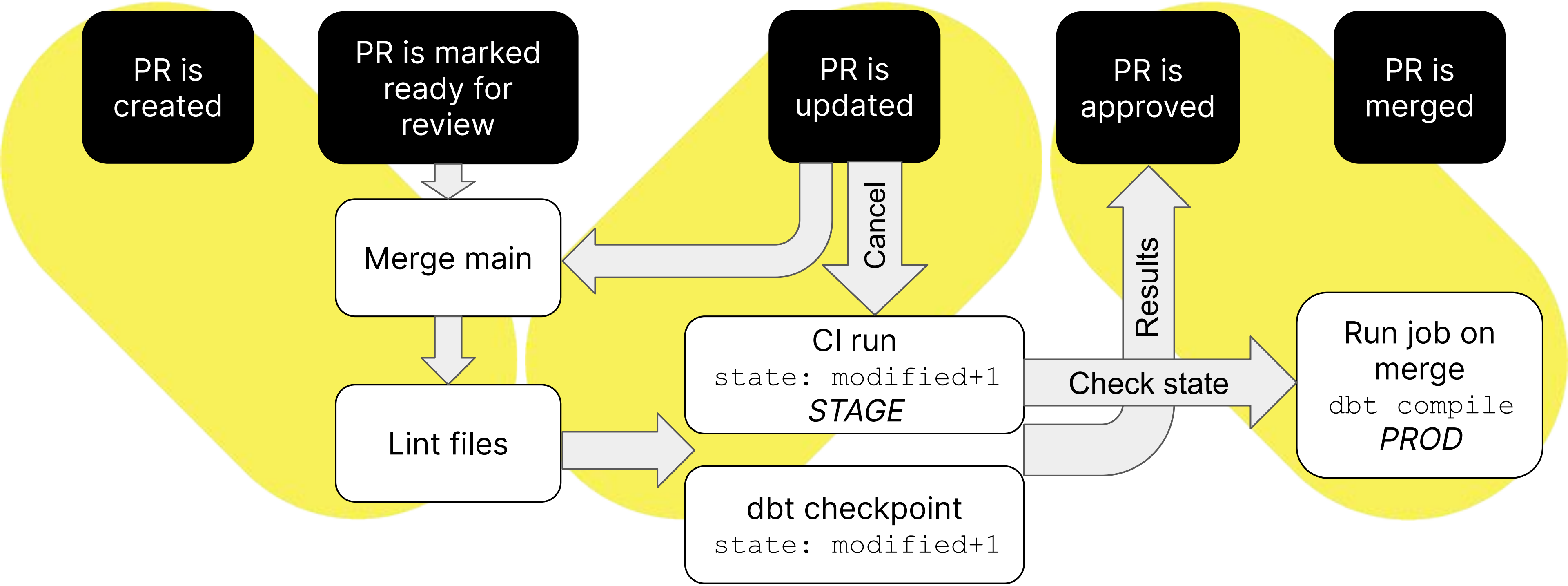
To be dropped: table PROJECTS.LIVESTREAM_PROMOTIONS__BASE_LIVESTREAM_ENGAGEMENTS

To be dropped: table PROJECTS.DSML_RANKING__USER_CATEGORY_FOLLOWS

To be dropped: table PROJECTS.DSML_METRICS__TIME_SPENT_BY_PANE

- Setup & access
  - Makefiles
  - SSO & IaC
- CI/CD
  - Pre-commit
  - Linting
  - Stage testing
- Code review
  - PR Checklists
  - CI test results
- Maintenance
  - Team tags for products/ processes
- Clean up

Case Study: CI/CD in the dbt repo



PR is created

PR is marked ready for review

PR is updated

PR is approved

PR is merged

Merge main

Lint files

Cancel

CI run
`state: modified+1`
STAGE

dbt checkpoint
`state: modified+1`

Results

Check state

Run job on merge
`dbt compile`
PROD

Future Directions

- Continuous deployment
- Modularity without dependency
- Query monitoring
    - Writing efficient SQL

Takeaways

- Everybody is welcome
- Guard rails not gates
- Keep lines short
- Keep rides running
- Have fun and be nice!