dbt Labs

Navigate and understand your data assets with dbt Explorer

April 12, 2024

Hey, I am Stephan and I work at dbt Labs. We make dbt

dbt Labs

dbt Cloud
used by >4,000 companies

dbt

80,000+ in dbt
Community Slack

Solutions Architect
Based in Berlin, Germany
Collecting football stadiums

Today's Agenda

1 **What is dbt Cloud** ⛅

Managing complexity at scale with dbt Cloud

2 **Why documenting and sharing organizational knowledge is hard** 🔍

How lack of visibility into data breeds more complexity

3 **How it works** 👷

dbt Explorer in practice and newest features

4 **Ok, show me** 👀

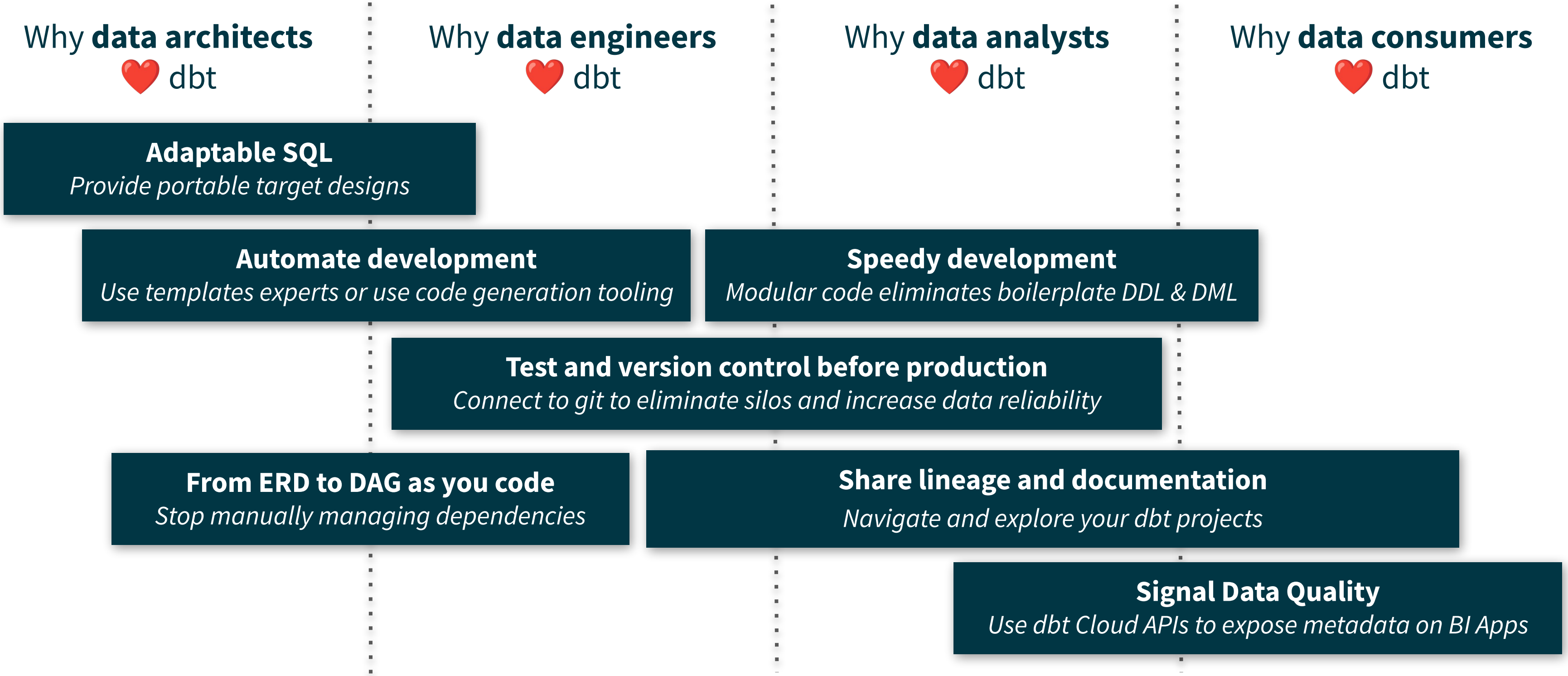
Demo time

5 **Q&A** 💬

Let's Discuss

Managing complexity at scale with dbt Cloud

dbt enables multiple roles to collaborate in one shared space



Why **data architects** ❤️ dbt

Why **data engineers** ❤️ dbt

Why **data analysts** ❤️ dbt

Why **data consumers** ❤️ dbt

Adaptable SQL
Provide portable target designs

Automate development
Use templates experts or use code generation tooling

Speedy development
Modular code eliminates boilerplate DDL & DML

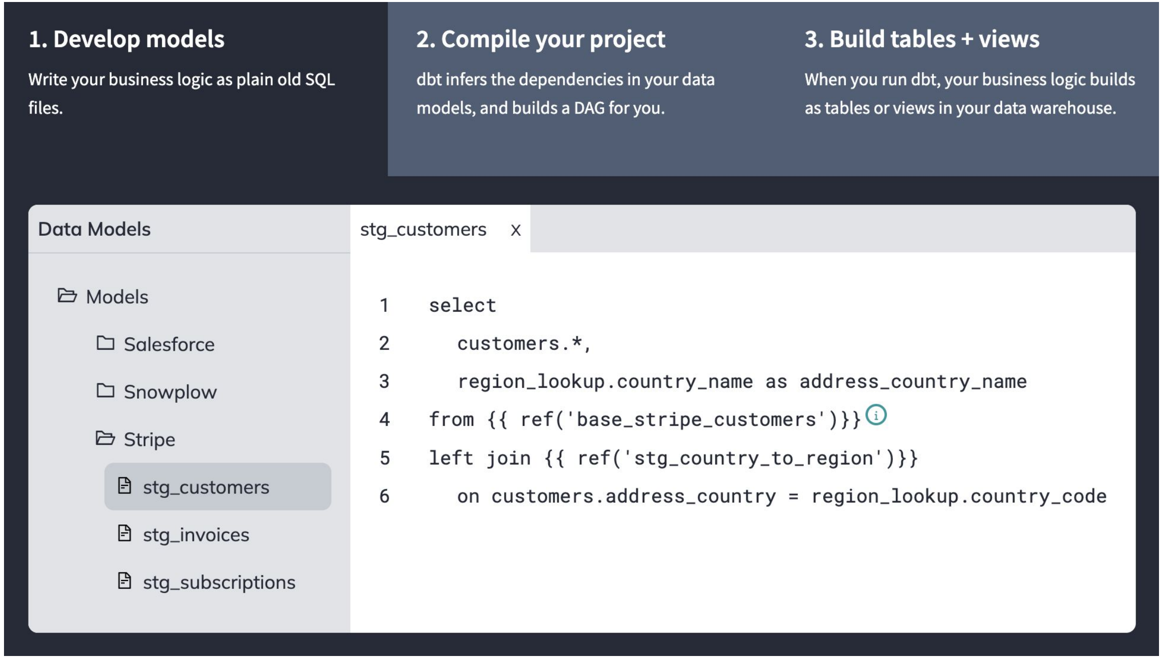
Test and version control before production
Connect to git to eliminate silos and increase data reliability

From ERD to DAG as you code
Stop manually managing dependencies

Share lineage and documentation
Navigate and explore your dbt projects

Signal Data Quality
Use dbt Cloud APIs to expose metadata on BI Apps

Step 1: Write reusable & referenceable logic with SQL + Jinja



1. Develop models

Write your business logic as plain old SQL files.

2. Compile your project

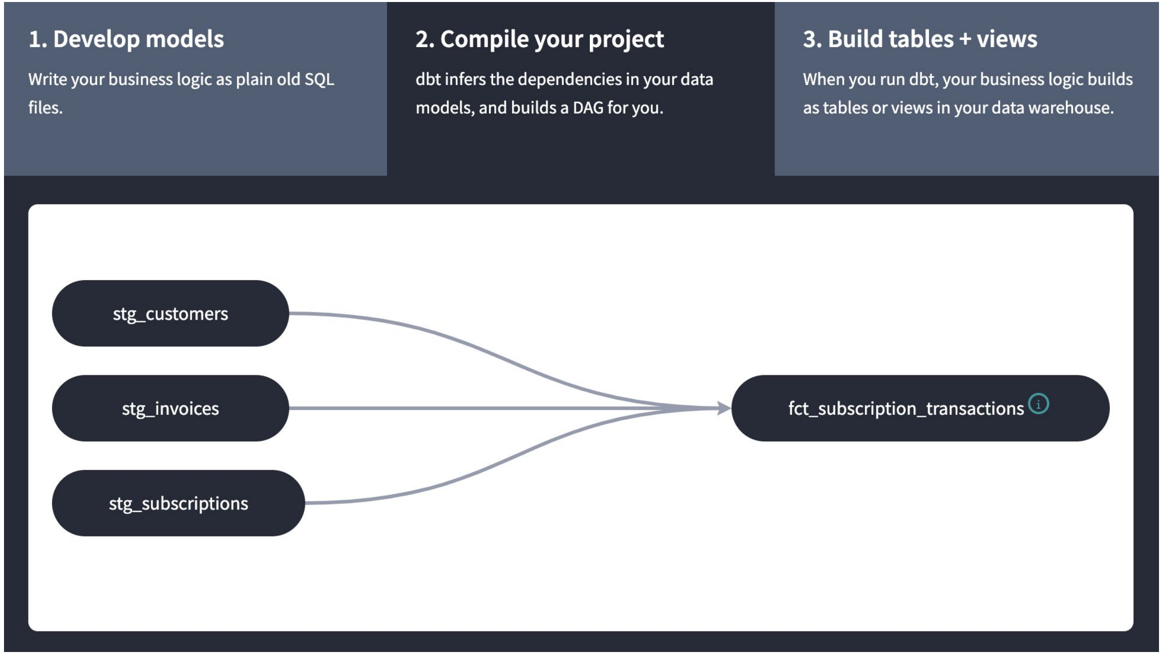
dbt infers the dependencies in your data models, and builds a DAG for you.

3. Build tables + views

When you run dbt, your business logic builds as tables or views in your data warehouse.

```sql
select
  customers.*,
  region_lookup.country_name as address_country_name
from {{ ref('base_stripe_customers')}}
left join {{ ref('stg_country_to_region')}}
  on customers.address_country = region_lookup.country_code
```

Step 2a: Infer lineage for automated dependency management.



1. Develop models

Write your business logic as plain old SQL files.

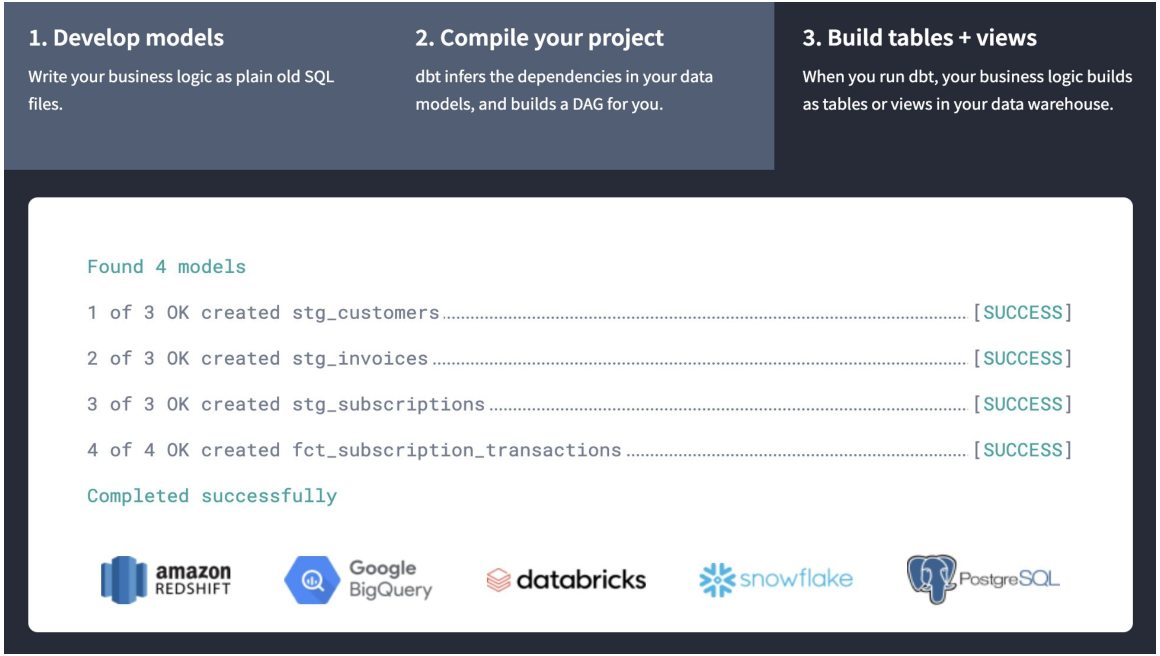
2. Compile your project

dbt infers the dependencies in your data models, and builds a DAG for you.

3. Build tables + views

When you run dbt, your business logic builds as tables or views in your data warehouse.

Step 2b: Run code to build datasets in your warehouse.



1. Develop models

Write your business logic as plain old SQL files.

2. Compile your project

dbt infers the dependencies in your data models, and builds a DAG for you.

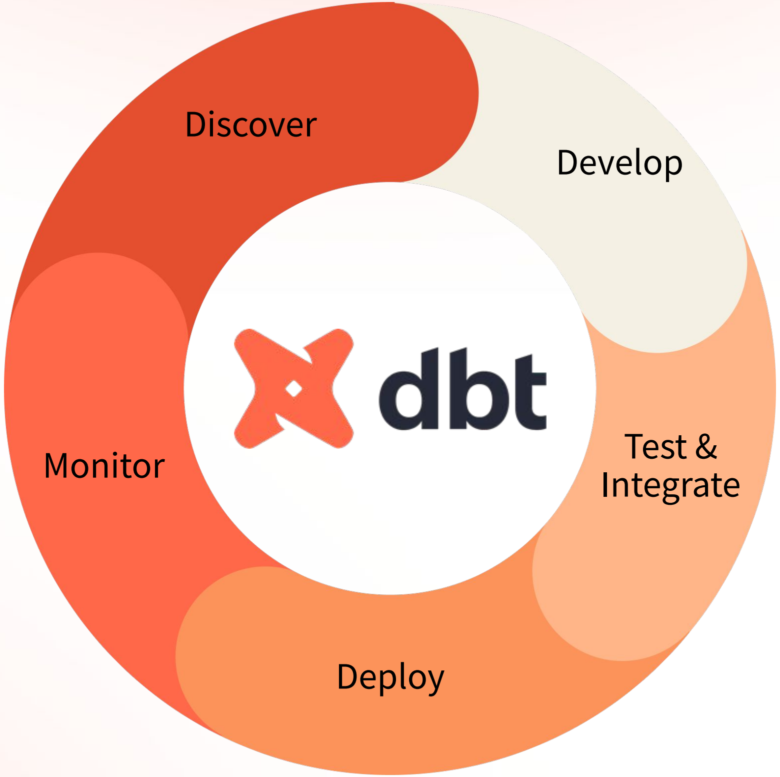
3. Build tables + views

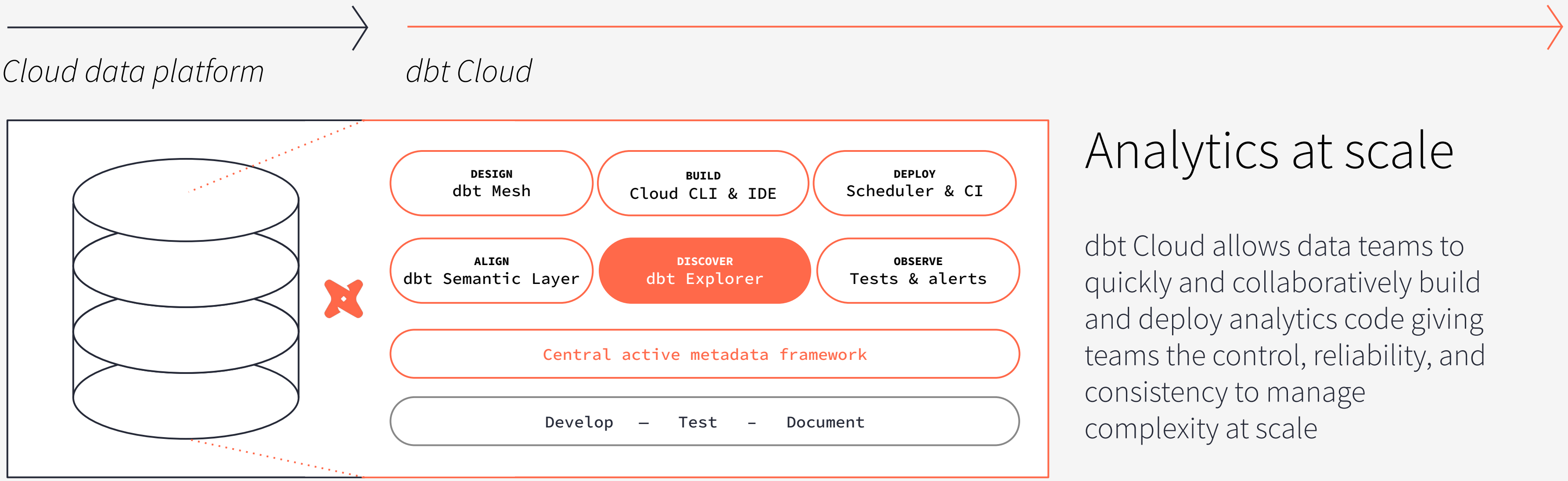
When you run dbt, your business logic builds as tables or views in your data warehouse.

```
Found 4 models
1 of 3 OK created stg_customers ....................................................... [SUCCESS]
2 of 3 OK created stg_invoices ........................................................ [SUCCESS]
3 of 3 OK created stg_subscriptions ................................................... [SUCCESS]
4 of 4 OK created fct_subscription_transactions ........................................ [SUCCESS]
Completed successfully
```

dbt Cloud enables the

Data Lifecycle





Analytics at scale

dbt Cloud allows data teams to quickly and collaboratively build and deploy analytics code giving teams the control, reliability, and consistency to manage complexity at scale

Data Contracts

A better way to discover data

Recap from Yesterday

“Document systems close to the source” -> Only integrate system that are documented

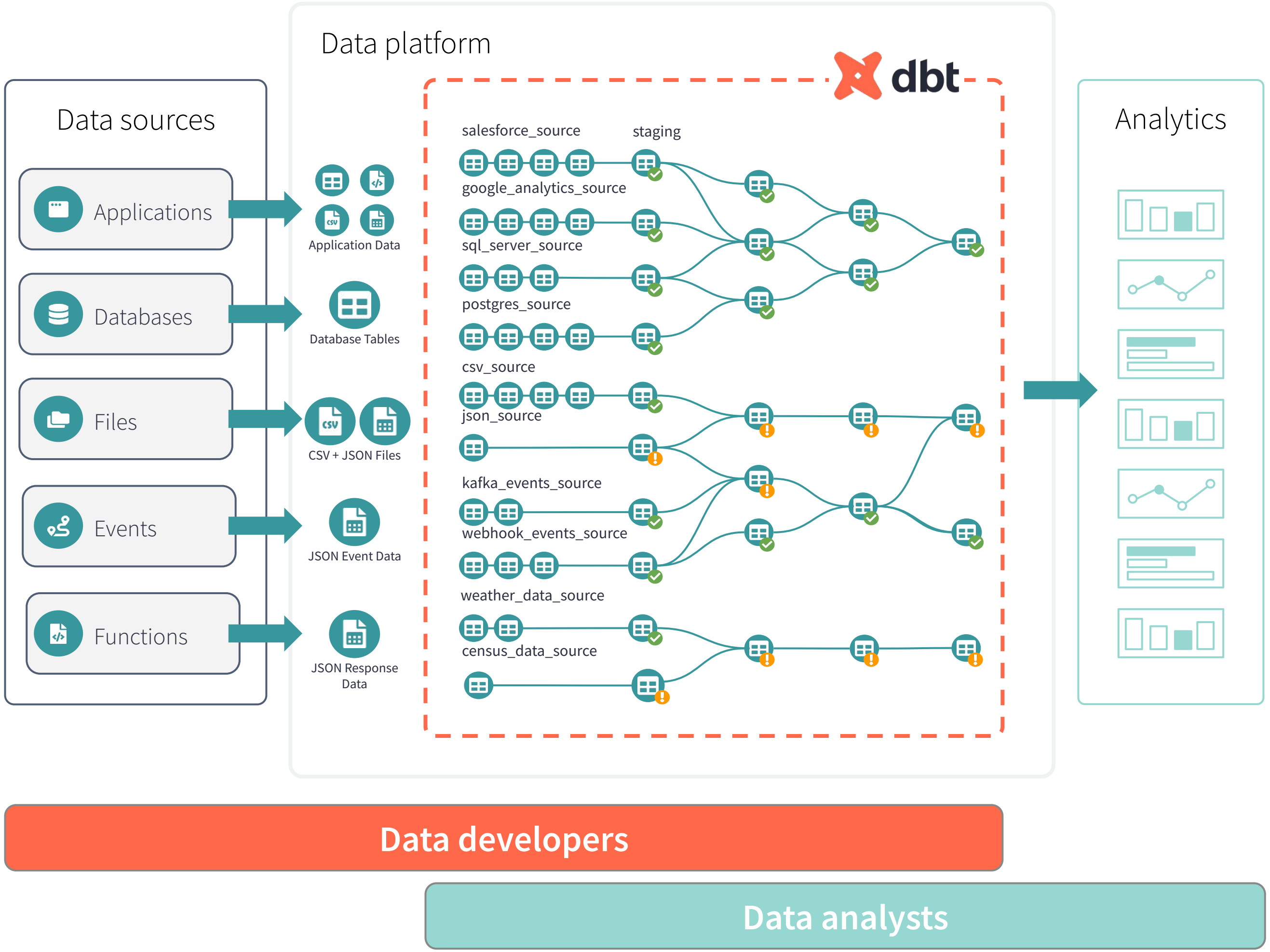
“Make metadata part of the data itself”

“Intentional Management and a structured processes”

“Federate metadata across different tools, but have a single place of truth”

“Make it easy for the business to collaborate over the metadata”

dbt Cloud is where developers and analysts come together to manage data at scale.

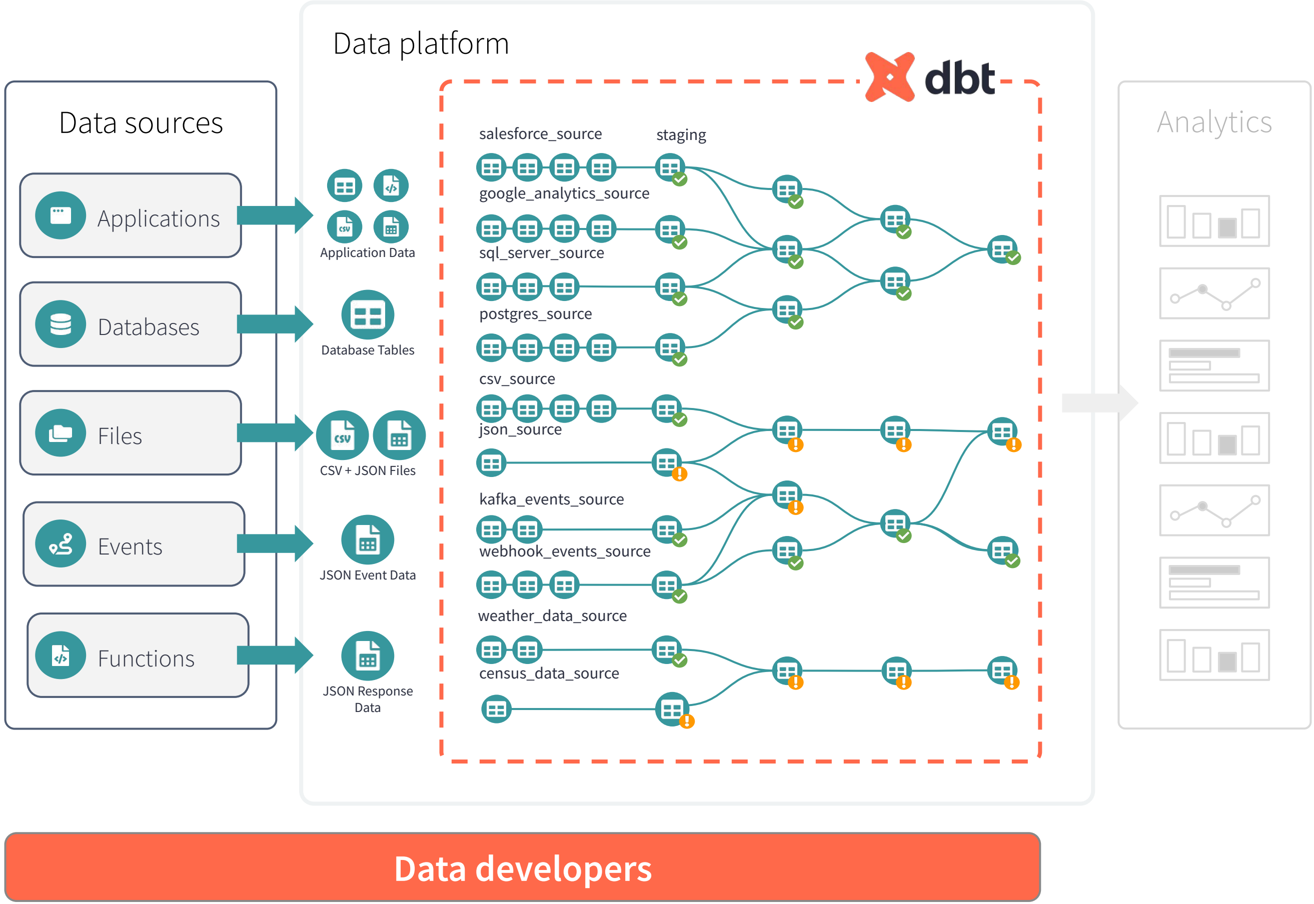


Data developers

Role: Build and maintain data pipelines for downstream analysts

Challenges at scale:

- Inefficient data development workstreams
- Diagnosing pipeline issues to ensure data quality and freshness
- **Maintaining governance with visibility into usage, ownership, and dependencies**
- Managing costs (infrastructure and time)

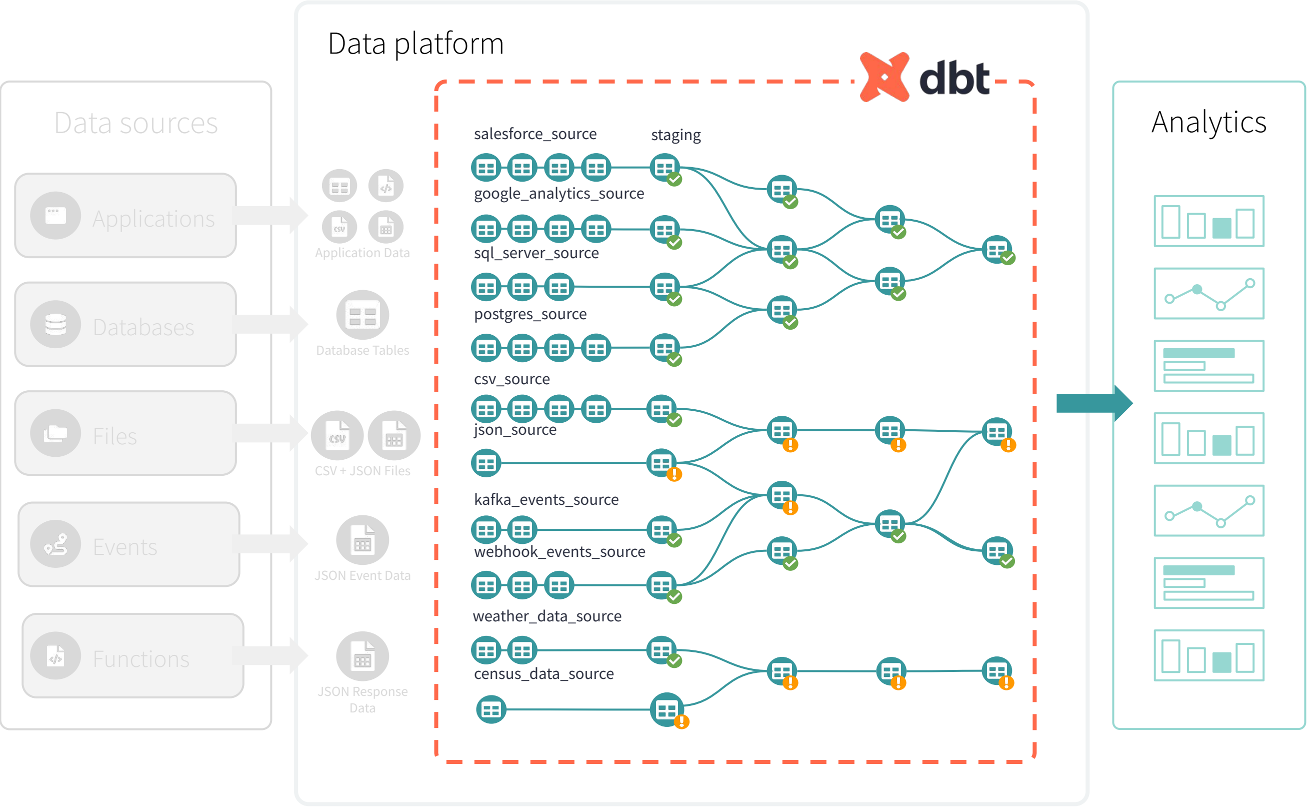


Data analysts

Role: Analyze data to inform business decisions and drive outcomes

Challenges at scale:

- **Understanding the meaning behind data products and how to use them**
- **Validating that the data can be trusted (freshness, testing)**
- Collaborating with data developers on real-time requests and modeling changes



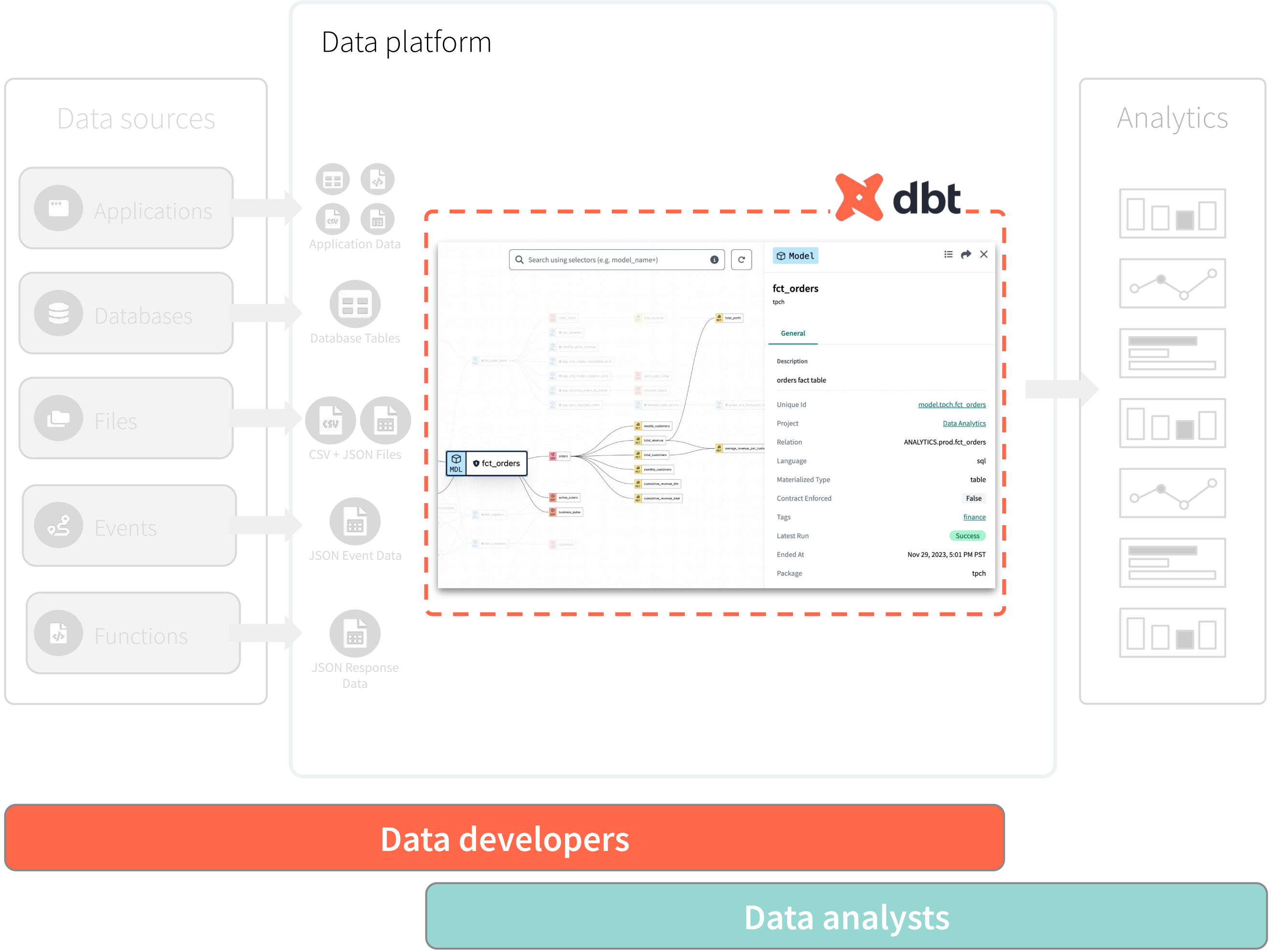
Data analysts

dbt Explorer

dbt Cloud’s knowledge base and interactive lineage graph

Data developers can understand data lineage so they can build, debug, and improve their dbt projects efficiently.

Data analysts can navigate, investigate, and leverage existing data products so they can deliver trustworthy insights.



fct_orders
tpch

Description
orders fact table

Unique Id	model.tpch.fct_orders
Project	Data Analytics
Relation	ANALYTICS.prod.fct_orders
Language	sql
Materialized Type	table
Contract Enforced	False
Tags	finance
Latest Run	Success
Ended At	Nov 29, 2023, 5:01 PM PST
Package	tpch

What about dbt Docs?

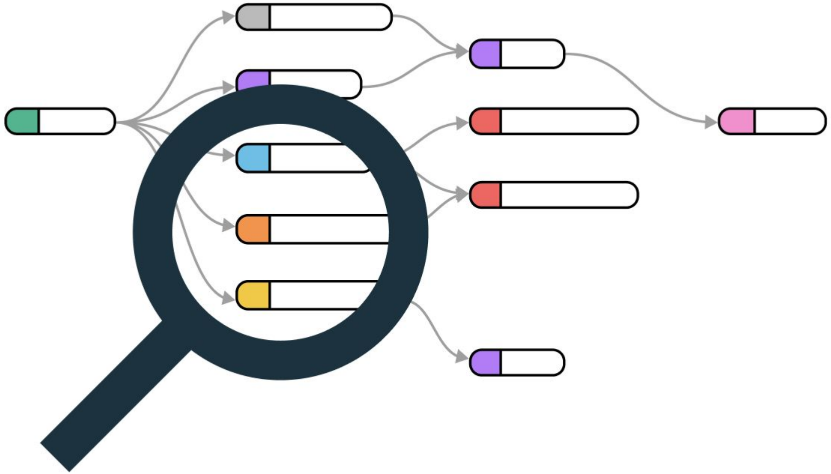
If you like dbt Docs…you’ll love dbt Explorer 😍

	Legacy dbt Docs	**dbt Explorer**
🏃 **Speed**	❌**Variable** Slows down, or doesn’t load for large dbt projects	✅ **Fast** Performant at scale, even for large projects
🤳 **Metadata**	❌**Static** Webpage with limited applied / run context	✅ **Live** Updates with rich metadata after each run
😎**Visibility**	❌**Limited** Only individual projects	✅ **Global** Multi-project lineage (dbt Mesh)

How it works

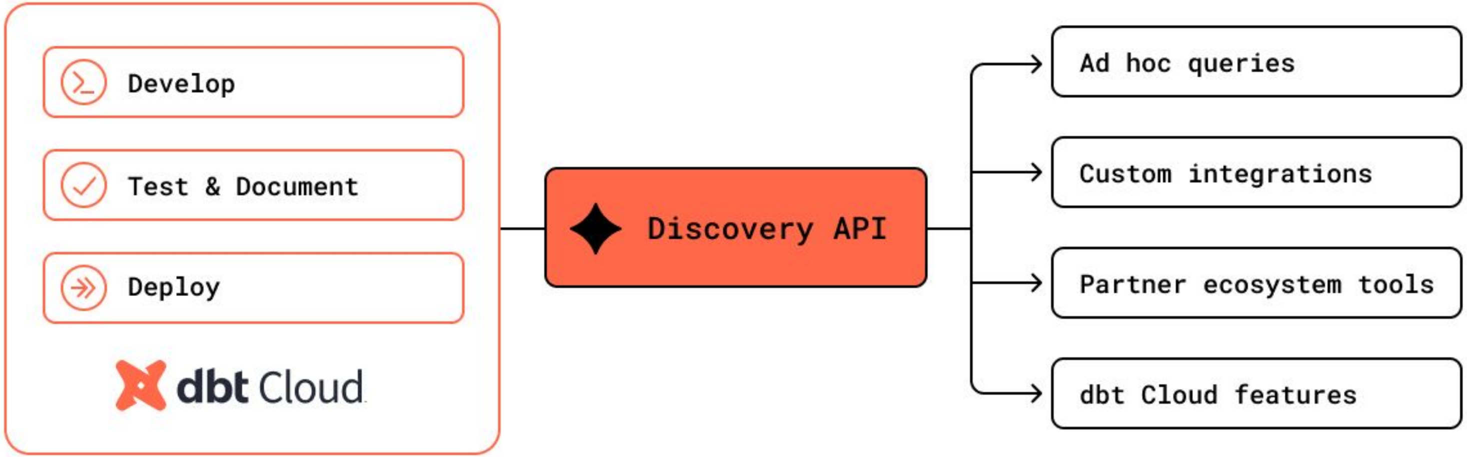
Discover Answers

dbt Cloud’s **metadata platform** powers experiences within dbt Cloud and its ecosystem to **improve data discovery, data quality, and pipeline performance**.



Ways of Using

The **Discovery API** is the gateway to the metadata platform. It enables users to **easily answer questions** using the results of dbt Cloud runs.

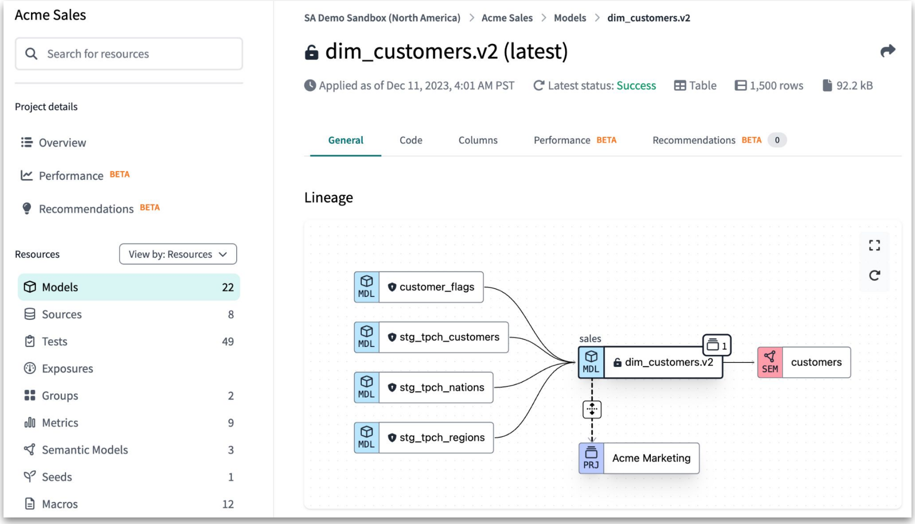


✅ **Public Preview**

dbt Cloud
Enterprise & Team Plan

dbt Explorer

Navigate and understand your dbt Cloud projects



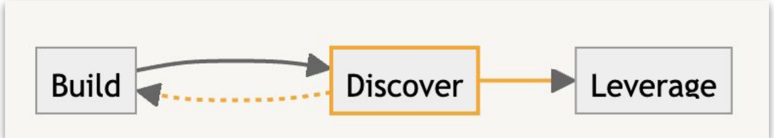
Acme Sales

SA Demo Sandbox (North America) > Acme Sales > Models > dim_customers.v2

dim_customers.v2 (latest)

Applied as of Dec 11, 2023, 4:01 AM PST   Latest status: Success   Table   1,500 rows   92.2 kB

Resources	
Models	22
Sources	8
Tests	49
Exposures	
Groups	2
Metrics	9
Semantic Models	3
Seeds	1
Macros	12



Build → Discover → Leverage

- **Discover & visualize existing assets**
  - ○ Birds’-eye-view of data state
  - ○ Find specific resources and their dependencies – reuse, don’t rebuild
  - ○ Understand quality and freshness via run metadata

- **Built for scale**
  - ○ Live, robust, and performant loading

- **Navigate global lineage (dbt Mesh)**
  - ○ Discover public models across projects

…but wait, there’s more!

✅ Public Preview

dbt Cloud
Enterprise & Team Plan

dbt Explorer

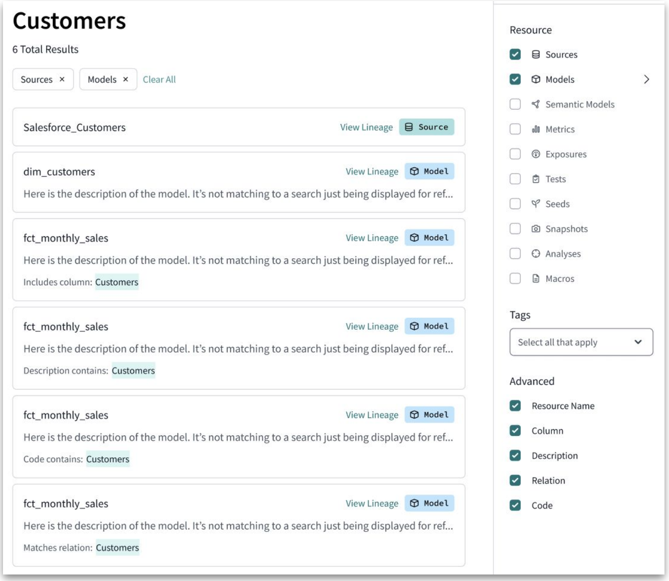
Improved search & lineage

Easily find resources and navigate projects

- Intuitive keyword search interface
  - Filter models
  - Search columns within resources

👀*Coming soon*

- Search descriptions, relation, and code
- Lineage "layers" to more easily distinguish nodes



Customers

6 Total Results

Sources × Models × Clear All

Salesforce_Customers — View Lineage — Source

dim_customers — View Lineage — Model
Here is the description of the model. It's not matching to a search just being displayed for ref...

fct_monthly_sales — View Lineage — Model
Here is the description of the model. It's not matching to a search just being displayed for ref...
Includes column: Customers

fct_monthly_sales — View Lineage — Model
Here is the description of the model. It's not matching to a search just being displayed for ref...
Description contains: Customers

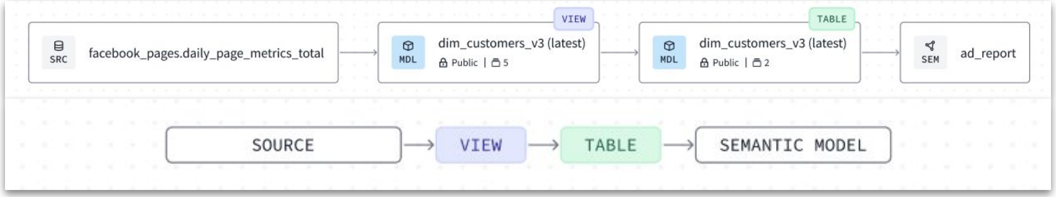
fct_monthly_sales — View Lineage — Model
Here is the description of the model. It's not matching to a search just being displayed for ref...
Code contains: Customers

fct_monthly_sales — View Lineage — Model
Here is the description of the model. It's not matching to a search just being displayed for ref...
Matches relation: Customers

Resource
☑ Sources
☑ Models
□ Semantic Models
□ Metrics
□ Exposures
□ Tests
□ Seeds
□ Snapshots
□ Analyses
□ Macros

Tags
Select all that apply

Advanced
☑ Resource Name
☑ Column
☑ Description
☑ Relation
☑ Code


SRC
facebook_pages.daily_page_metrics_total
VIEW
MDL
dim_customers_v3 (latest)
Public | 5
TABLE
MDL
dim_customers_v3 (latest)
Public | 2
SEM
ad_report
SOURCE
VIEW
TABLE
SEMANTIC MODEL

✅ Public Preview

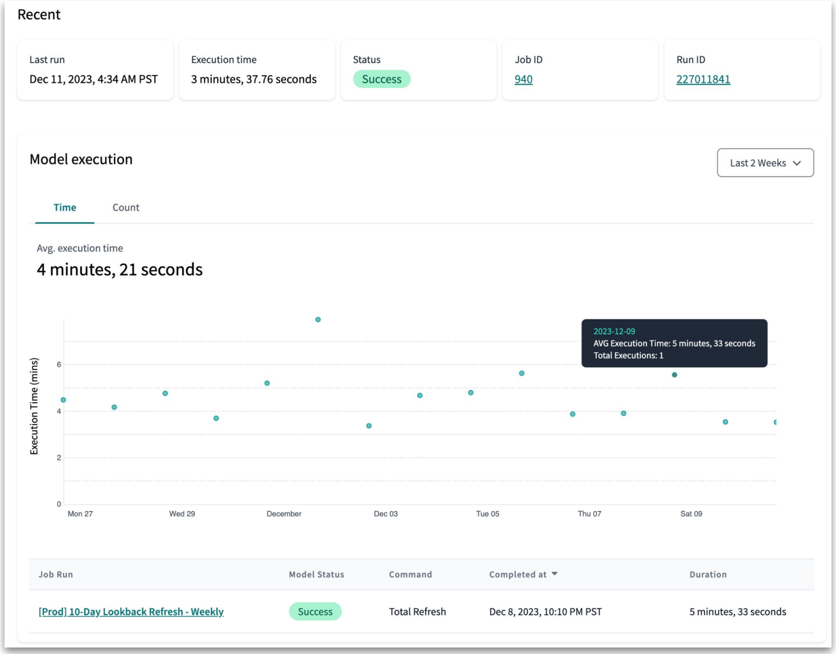
dbt Cloud
Enterprise & Team Plan

dbt Explorer

Model performance analysis

Discover low-hanging fruit for optimization

- Analyze historical trends of model execution time, count, and tests
- Pinpoint long-running or often-failing models
- Quickly identify opportunities to reduce infrastructure costs, improve reliability, and save data team time



Recent

Last run	Execution time	Status	Job ID	Run ID
Dec 11, 2023, 4:34 AM PST	3 minutes, 37.76 seconds	Success	940	227011841

Model execution

Last 2 Weeks

Avg. execution time

4 minutes, 21 seconds

2023-12-09
AVG Execution Time: 5 minutes, 33 seconds
Total Executions: 1

Job Run	Model Status	Command	Completed at ▼	Duration
[Prod] 10-Day Lookback Refresh - Weekly	Success	Total Refresh	Dec 8, 2023, 10:10 PM PST	5 minutes, 33 seconds

✅ **Public Preview**

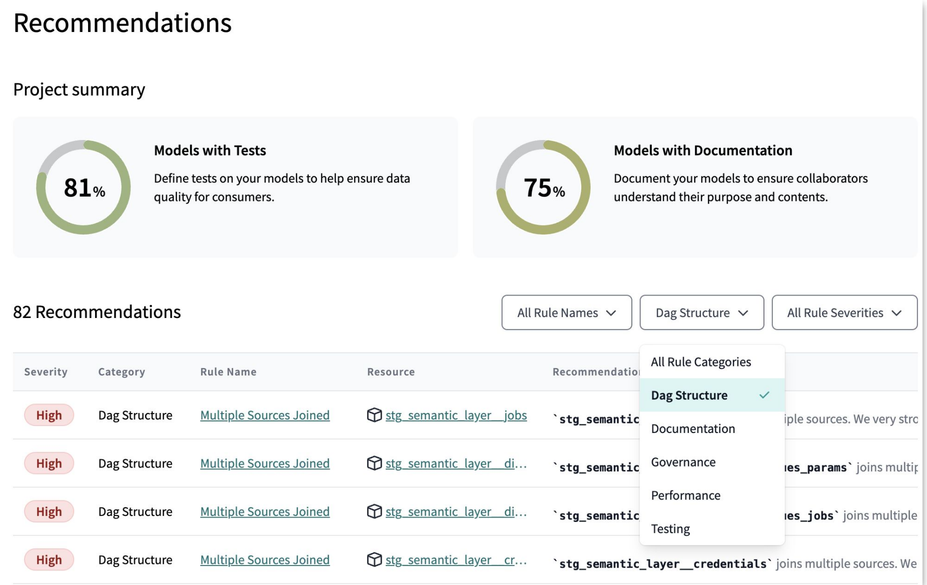
dbt Cloud
Enterprise & Team Plan

dbt Explorer

Project recommendations

Proactively improve your data pipelines using dbt Cloud metadata

- Quickly understand test and documentation coverage and project health at a glance
- Tackle dbt project improvements efficiently with guidance on how and where to resolve issues
- Recommendations powered by dbt Cloud metadata and best practices



Recommendations

Project summary

81% **Models with Tests**
Define tests on your models to help ensure data quality for consumers.

75% **Models with Documentation**
Document your models to ensure collaborators understand their purpose and contents.

82 Recommendations

Severity	Category	Rule Name	Resource	Recommendatio
High	Dag Structure	Multiple Sources Joined	stg_semantic_layer__jobs	`stg_semantic … iple sources. We very stro
High	Dag Structure	Multiple Sources Joined	stg_semantic_layer__di…	`stg_semantic … ies_params` joins multip
High	Dag Structure	Multiple Sources Joined	stg_semantic_layer__di…	`stg_semantic … ies_jobs` joins multiple
High	Dag Structure	Multiple Sources Joined	stg_semantic_layer__cr…	`stg_semantic_layer__credentials` joins multiple sources. We

✅ **Public Beta**

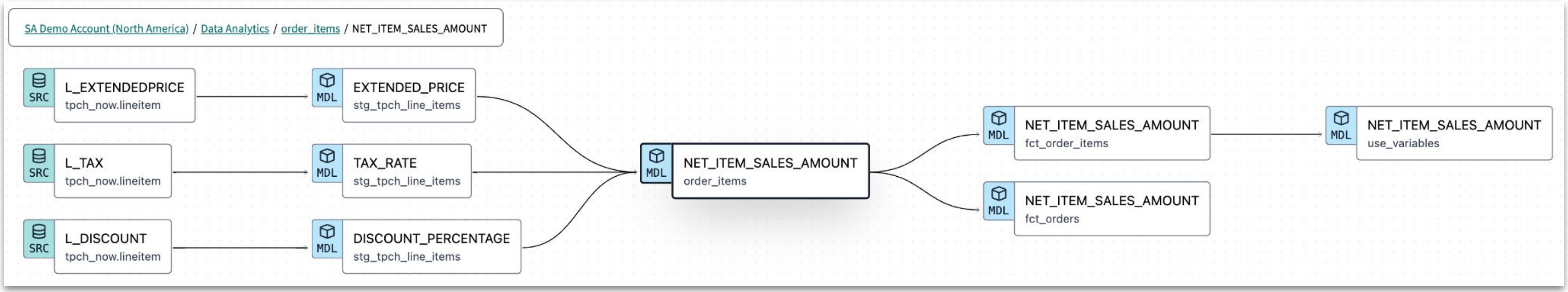
dbt Cloud
Enterprise Plan

dbt Explorer

Column-level lineage

Identify and resolve data issues faster

- **Audit**: Visualize how data moves through and is used in your dbt project
- **Root cause**: Improve time to detect and resolve data quality issues, tracking back to the source
- **Impact analysis**: Trace transformations via a detailed view into column dependencies


SA Demo Account (North America) / Data Analytics / order_items / NET_ITEM_SALES_AMOUNT
SRC L_EXTENDEDPRICE tpch_now.lineitem
SRC L_TAX tpch_now.lineitem
SRC L_DISCOUNT tpch_now.lineitem
MDL EXTENDED_PRICE stg_tpch_line_items
MDL TAX_RATE stg_tpch_line_items
MDL DISCOUNT_PERCENTAGE stg_tpch_line_items
MDL NET_ITEM_SALES_AMOUNT order_items
MDL NET_ITEM_SALES_AMOUNT fct_order_items
MDL NET_ITEM_SALES_AMOUNT fct_orders
MDL NET_ITEM_SALES_AMOUNT use_variables

dbt Explorer benefits:

- **Stateful:** Monitor data flows with applied / run metadata and observability built-in
- **Visibility:** Elevate data lineage as a first-class citizen
- **Performant:** Reliable and consistently available at enterprise scale—within and across dbt projects (dbt Mesh)
- **Integrated:** Use cross-platform metadata and dbt workflows to continuously improve user experience

dbt

Thanks for joining!

Click “Docs” in Goldcast to access these links!

✓ **Get started today with your own dbt Cloud trial!**

Want to chat with a dbt specialist for all of your nitty-gritty questions?
You can do that anytime here.