

dbt: Transforming the Value of Boston's Data



Lunch & Learn

Presented on 2 November 2023

Presented by Jenna Jordan, Analytics Team Data Engineer

CITY OF BOSTON



Overview

Current state of data management

Introducing dbt

Key dbt features solve our pain points

Demo of dbt generated Data Catalog

Project Timeline

Next Steps

Appendix

dbt CAN HELP SOLVE MANY CURRENT PAIN POINTS

GOAL

CURRENT STATE

Data Catalog

- Not in a central/consistent location
- Not always accessible or self-service
- Inconsistent format & quality

Data Governance

- Data lineage, ownership, downstream use, and expected freshness are unclear and usually not documented

Data Quality

- DUT failures not accessible/transparent

Change Enablement

- Data warehouse is a black box for those not on Analytics team

Faster Outcomes

- Always seeking continuous improvement

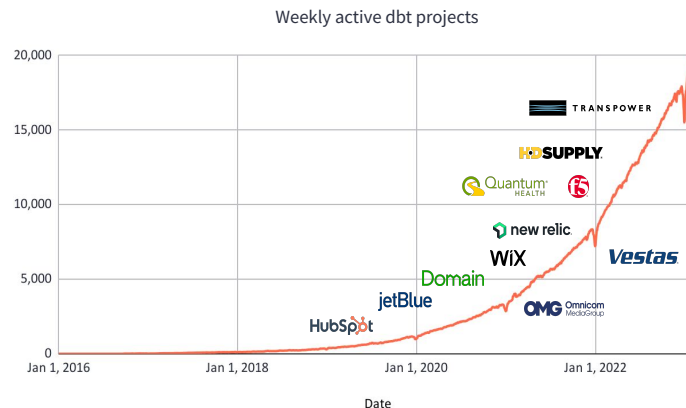
HOW WE GET THERE - INTRODUCING dbt



dbt (Data Build Tool) is a:

“**tool that enables data teams to transform data using analytics engineering best practices**”

- Popular (over 30,000 companies)
- Mature (5+ years old)
- Free (open source python package)
- Actively maintained/developed & documented by dbt Labs
- Active community of users & developers (over 90,000 people)
- Foundational tool in the modern data stack



Goal

dbt Feature

Data Catalog

- Automatically generated & updated data catalog website

Data Governance

- Data lineage is automatically documented & visualized
- Data Ownership explicitly documented in 1 place

Data Quality

- Easier to add tests & create custom tests
- Test failures recorded in database
- Separate dev environment

Change Enablement

- Automatic dependency graphs, including Exposures (external dependencies) on data catalog site

Faster Outcomes

- Packages & macros (enable DRY code)
- Declarative (dbt handles execution)

And much more!

COMPREHENSIVE DOCUMENTATION IS ESSENTIAL FOR DATA GOVERNANCE

Overview

Project Database
Group

Sources

- internal_stg
- open_stg
- restricted_stg

Exposures

- Analysis
- Application
- Dashboard

Projects

- analytics

Analytics Data Catalog

Welcome to the Analytics Data Catalog, where you can find documentation on the data assets in the Analytics Team's Data Warehouse. If you know what you are looking for, you can start typing in the search bar at the top. If you would like to explore the catalog, navigate using the left sidepane. If at any time you would like to view the lineage graph (for everything, or for the specific resource you are viewing), click the circular red button on the bottom right of the page.

Please note that this documentation site is a work in progress as we migrate our data assets over to a new set of schemas and our existing ETL infrastructure over to dbt.

Navigation

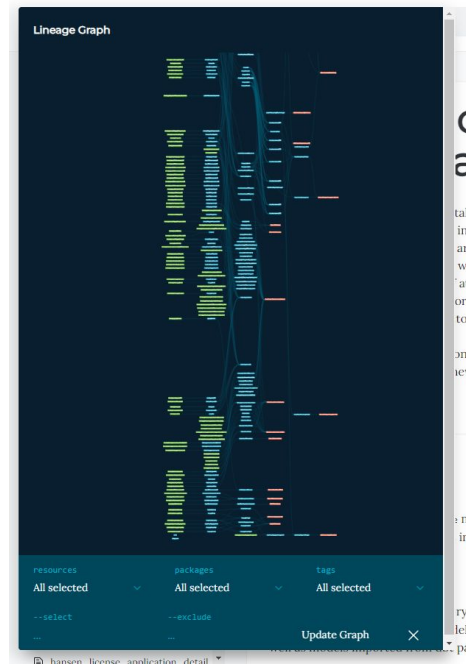
You can use the **Project** and **Database** navigation tabs on the left side of the window to explore the models in this dbt project.

Project Tab

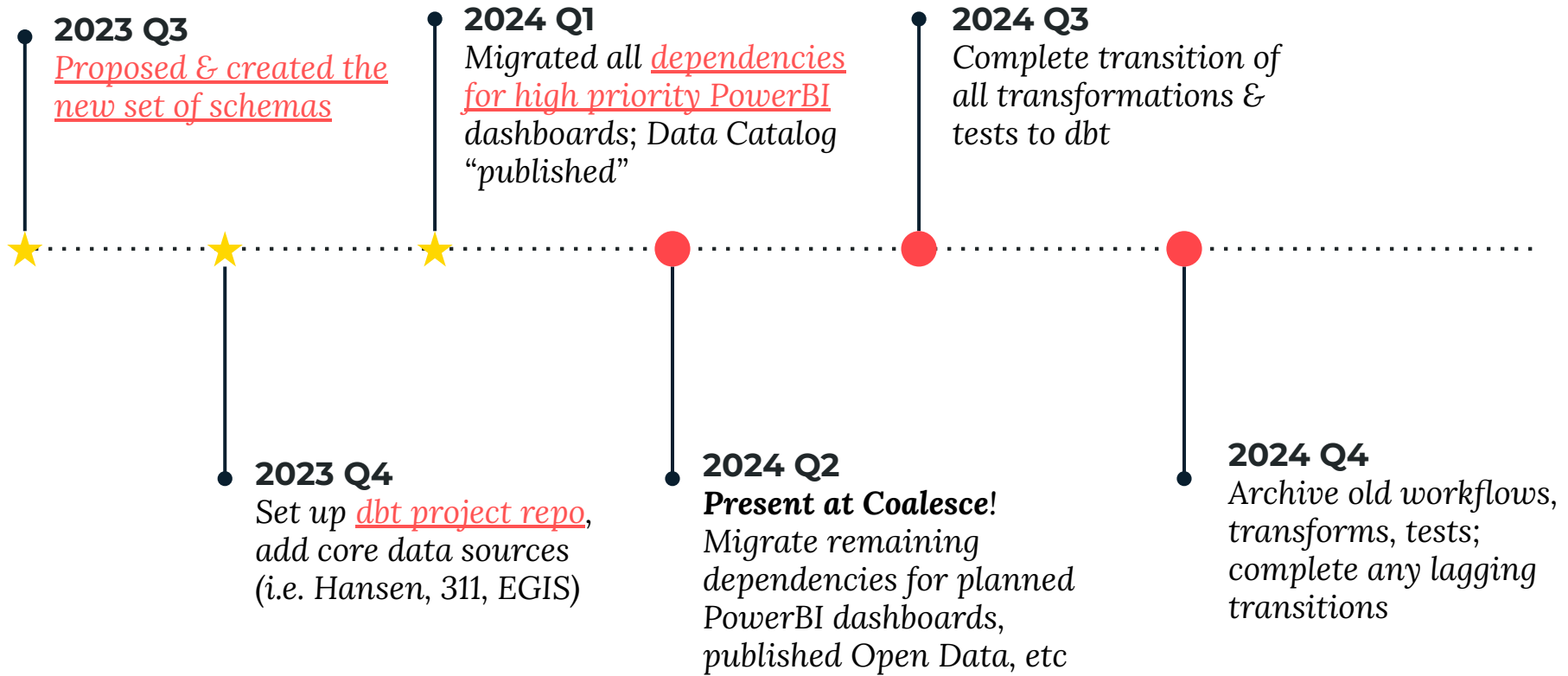
The **Project** tab mirrors the directory structure of the dbt project. In this tab, you can see all of the models defined in the dbt project, as well as models imported from dbt packages.

Sources

You can see documentation for all Source tables under the Sources header, divided according to data access level (open, internal, and restricted).



dbt PROJECT TIMELINE IS ON TRACK



Sharing out our work with dbt

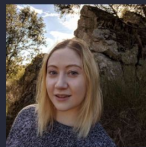
coalesce^o

Tuesday, Oct 17 2:30 PM PDT

From coast to coast:
Implementing dbt in the public
sector



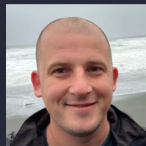
San Diego
Crafted by  dbt Labs.



Jennifer Jordan

Data Engineer, City of Boston,
Department of Innovation & Technology

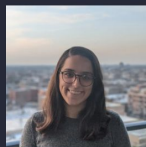
she/her



Ian Rose

Senior Data Engineer, California
Office of Data and Innovation

he/him



Laurie Merrell

Senior Analytics Engineer,
Jarvus Innovations

she/her

Recording now available
on [YouTube!](#)

[Repository](#) with
supporting resources



WHAT POTENTIAL DO YOU SEE IN THIS TOOL FOR YOUR TEAM?

Do you have any questions about dbt or the analytics data warehouse?

Do you think you would use the Data Catalog?

How would this resource be most helpful to your work?

Do you have any documentation on your data sources that could be added to the data catalog site?

Are you interested in directly contributing to the GitHub repo?

Next Steps

- Getting a boston.gov domain name for the data catalog
- Integrating with Access Boston for login
- Opening the data catalog up to all city workers after a security review

THANK YOU

