

---

# DuckDBT: Not a database or a dbt adapter but a secret third thing

---

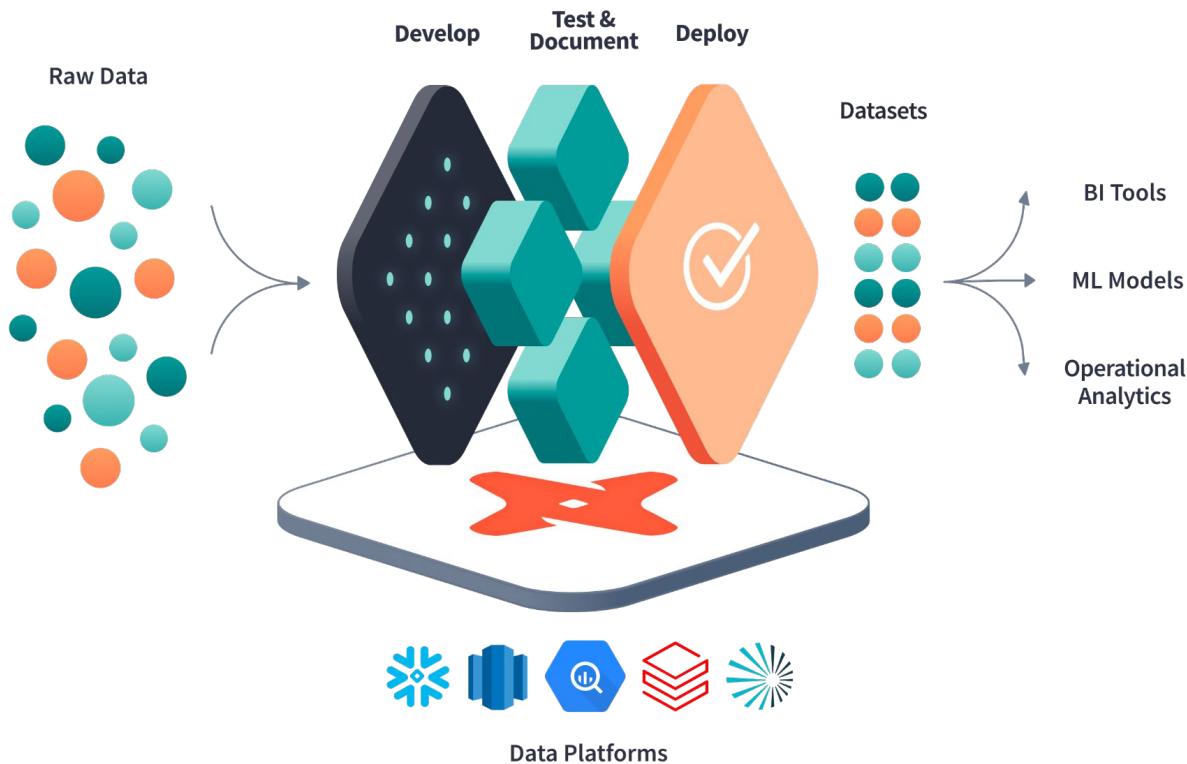
Josh Wills, DuckDB Enthusiast

# Who, exactly, am I?

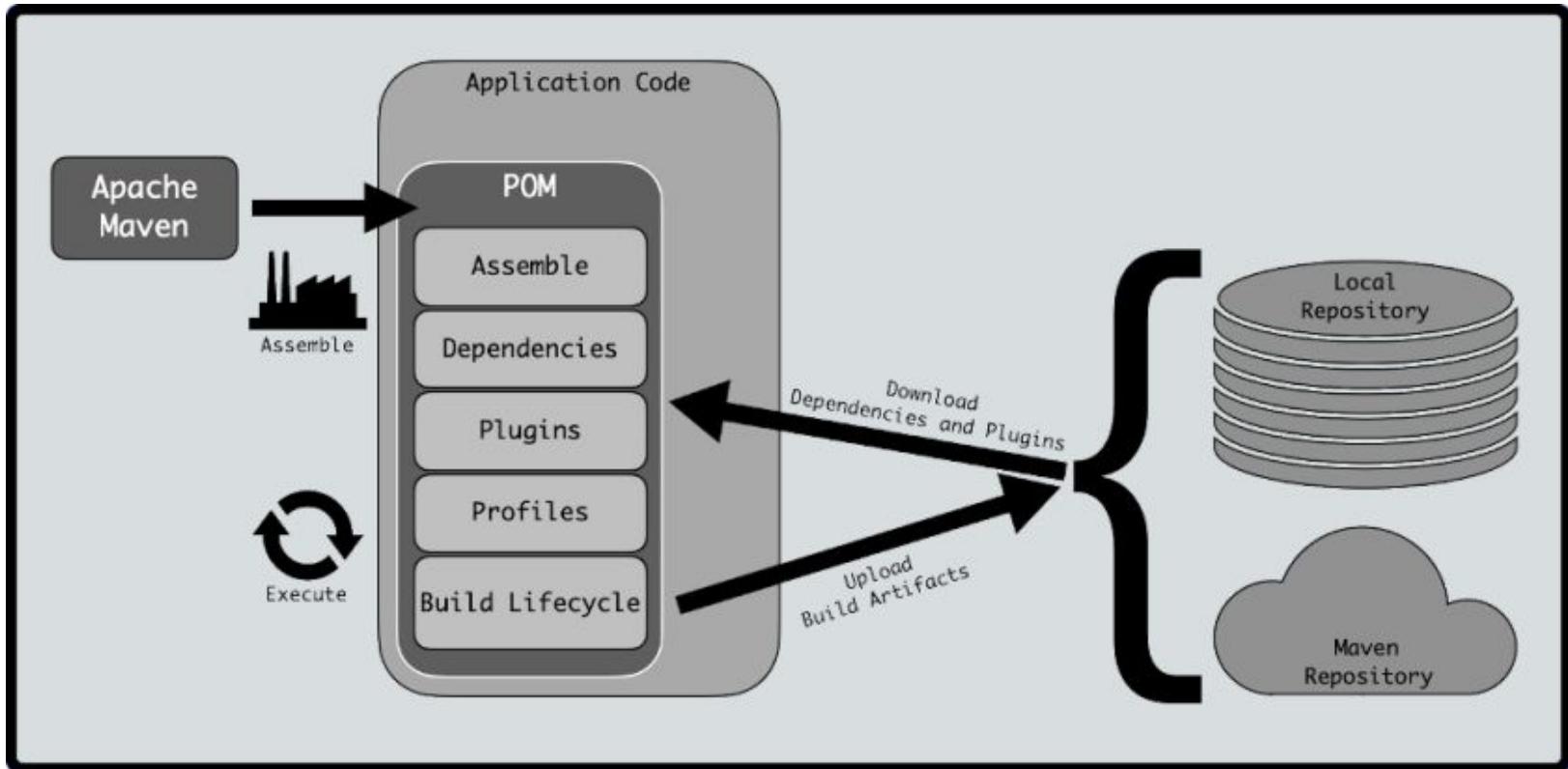
- Google Engineer  
(2007-11)
- Cloudera Data Science  
(2011-15)
- Slack Data and Search  
(2015-2019)
- WeaveGrid Engineer  
(2020 - 2022)



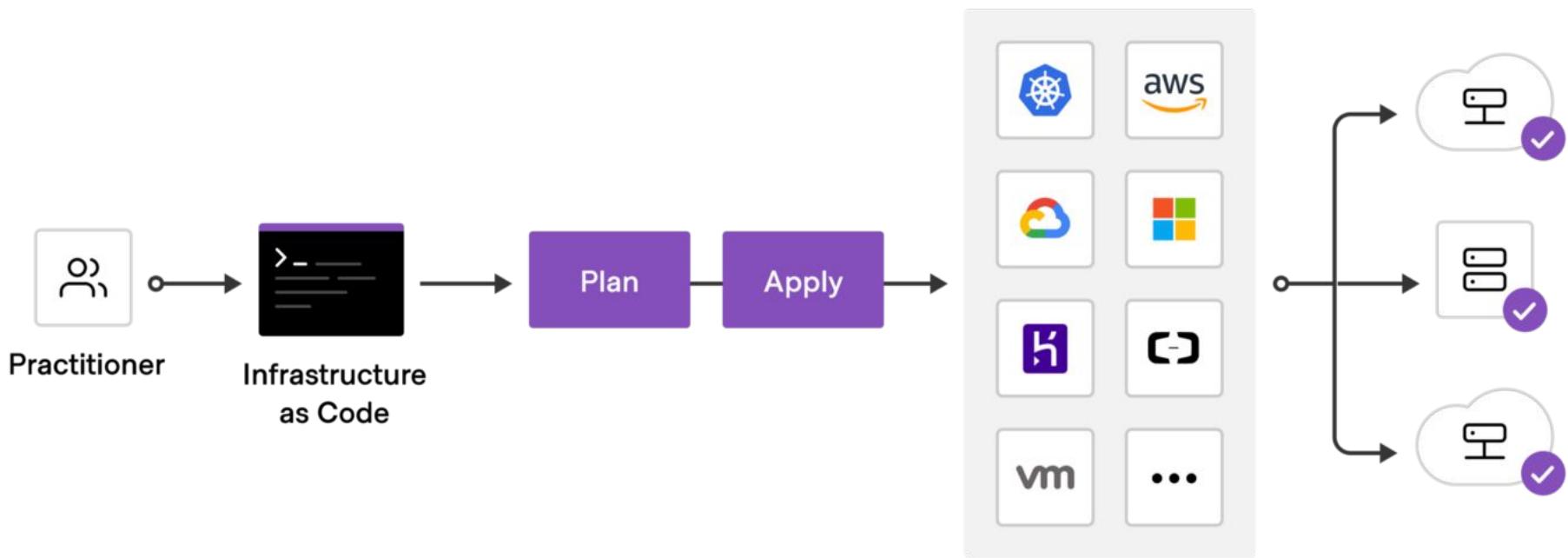
# What, exactly, is dbt?



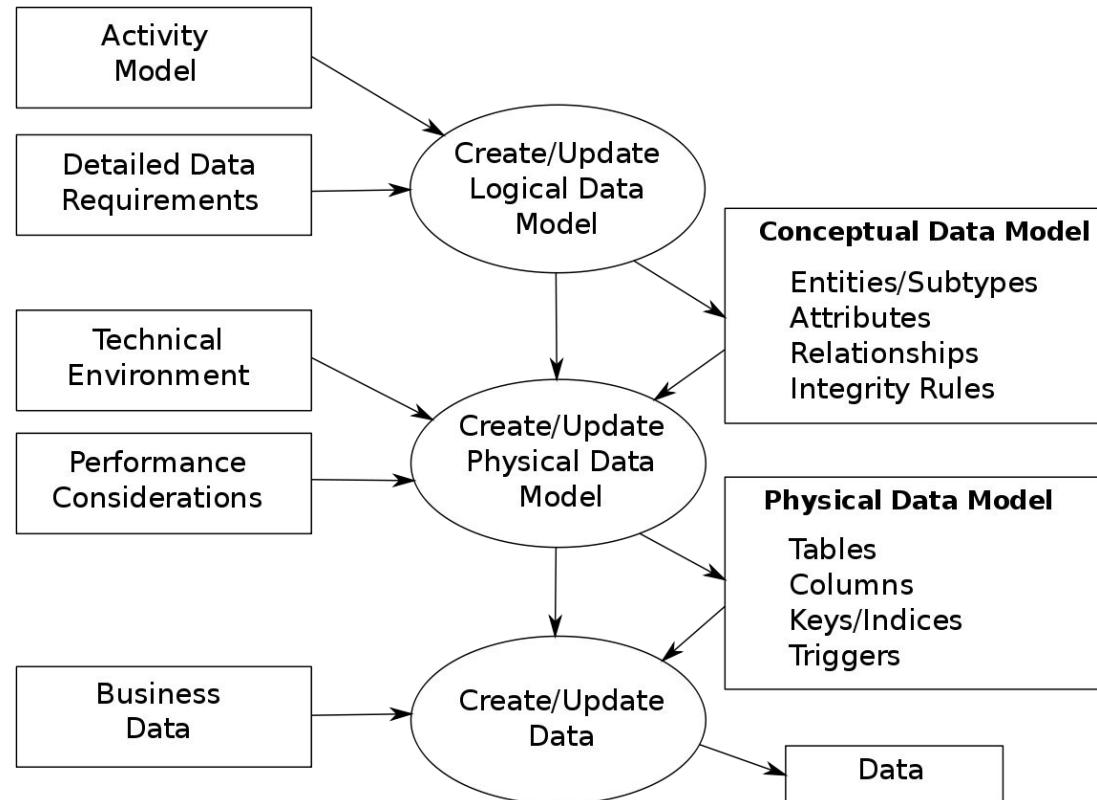
# The Data BUILD Tool For Analytics Engineers



# Like Terraform For Analytical Data Models



# Data Modeling Without The Data Warehouse



# The dbt-duckdb Vision

📌 Pinned by you



**Josh Wills ( /shrug )** 10:07 AM

The dbt-duckdb Vision:

1. Be the absolute easiest and fastest way for *anyone* to get up and running as an analytics engineer and support the ecosystem by e.g. helping DX teams to put together demos and supporting CICD for package developers 🇬🇧 😊
2. Push the limits of dbt as a tool and make it do all of the things that dbt Labs doesn't want you to be able to do with it. 🏴‍☠️ 😈



14



1



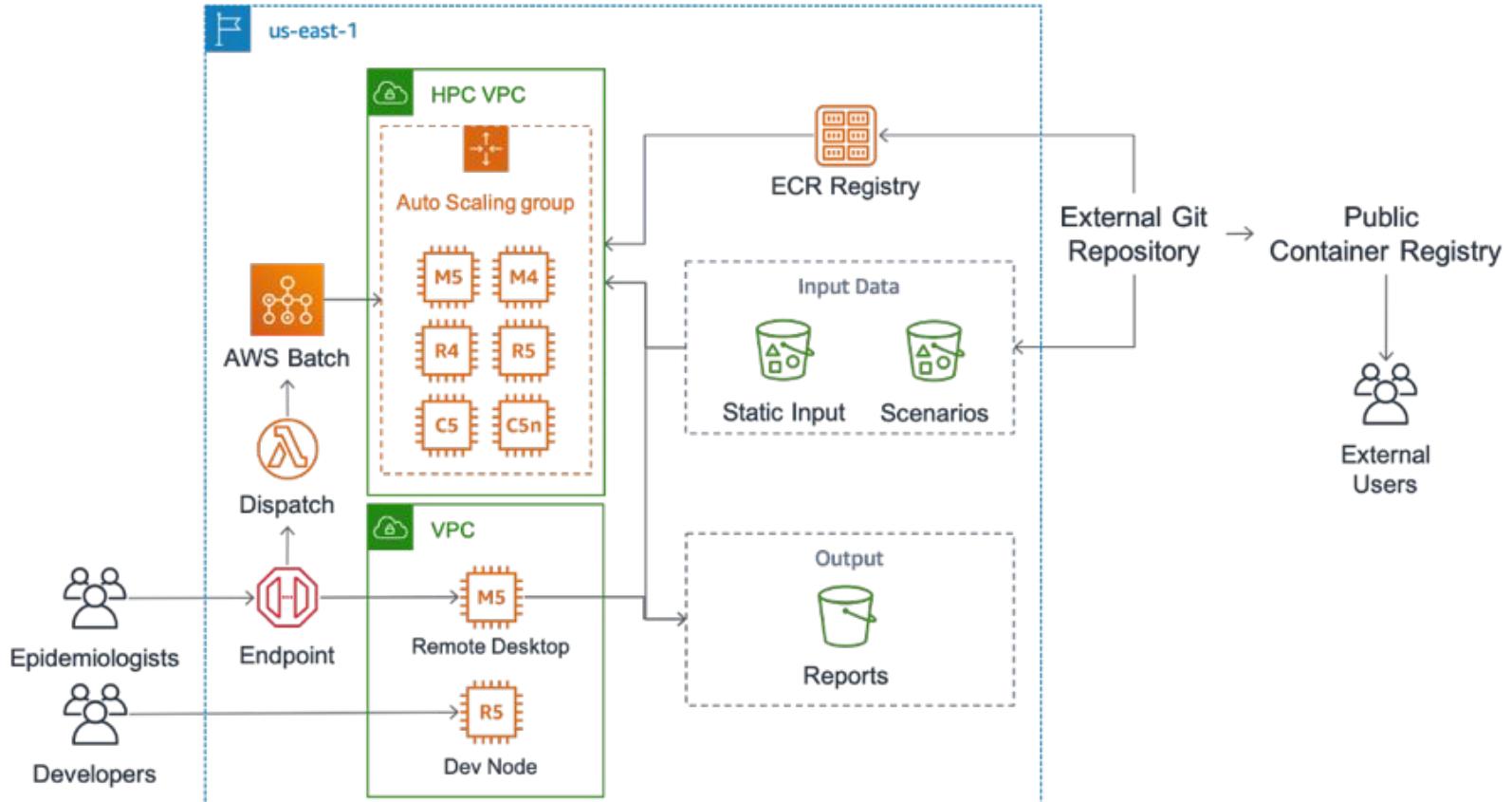
3



# My Super Villain Origin Story



# March 2020



# In the beginning... 🇬🇧

Initial commit of dbt-duckdb

master v1.3.1 ... 1.4.0

jwills committed on Sep 9, 2020 0 parents commit b6ec4d8

Showing 9 changed files with 241 additions and 0 deletions.

Filter changed files

78 .gitignore

12 README.md

12 dbt/adapters/duckdb/\_\_init\_\_.py

12 dbt/adapters/duckdb/\_init\_\_.py

```
@@ -0,0 +1,12 @@
+ ## dbt-duckdb
+
+ ### Installation
+ This plugin can be installed via pip:
+ ``
+ $ pip install dbt-duckdb
+ ``
+
+ ### Configuring your profile
+ [DuckDB] (http://duckdb) is an embedded database, similar to SQLite, but designed for OLAP-style analytics instead of OLTP. The only configuration parameter
+ for the profile is the `database` which should be a path to a local file that will contain the DuckDB database.
```

# My First Pull Request To DuckDB

## Table functions information\_schema\_schemata() and information\_schema\_tables() #899

Edit < Code ▾

Merged Mytherin merged 11 commits into `duckdb:master` from `jwills:jwills_info_schema` on Sep 14, 2020

Conversation 18

Commits 11

Checks 0

Files changed 9

+358 -1



jwills commented on Sep 11, 2020

Contributor ...

Per issue #658, we would like to have support for the metadata tables defined in the `information_schema` schema (specifically, `schemata`, `tables`, and `columns`) to support tools like DBeaver and dbt that depend on this metadata in order to operate against a broad class of SQL databases. This PR adds the `information_schema_schemata()` and `information_schema_tables()` table functions that implement the columns in `information_schema.schemata` and `information_schema.tables` in support of this goal.

Since this is my first PR to the project (and first time writing C++ in a pretty long time), I thought I would submit this bit to get feedback on while I work on implementing the more involved (but still straightforward) `information_schema_columns()` table function in a separate PR.



Reviewers

hannes



Mytherin



Assignees

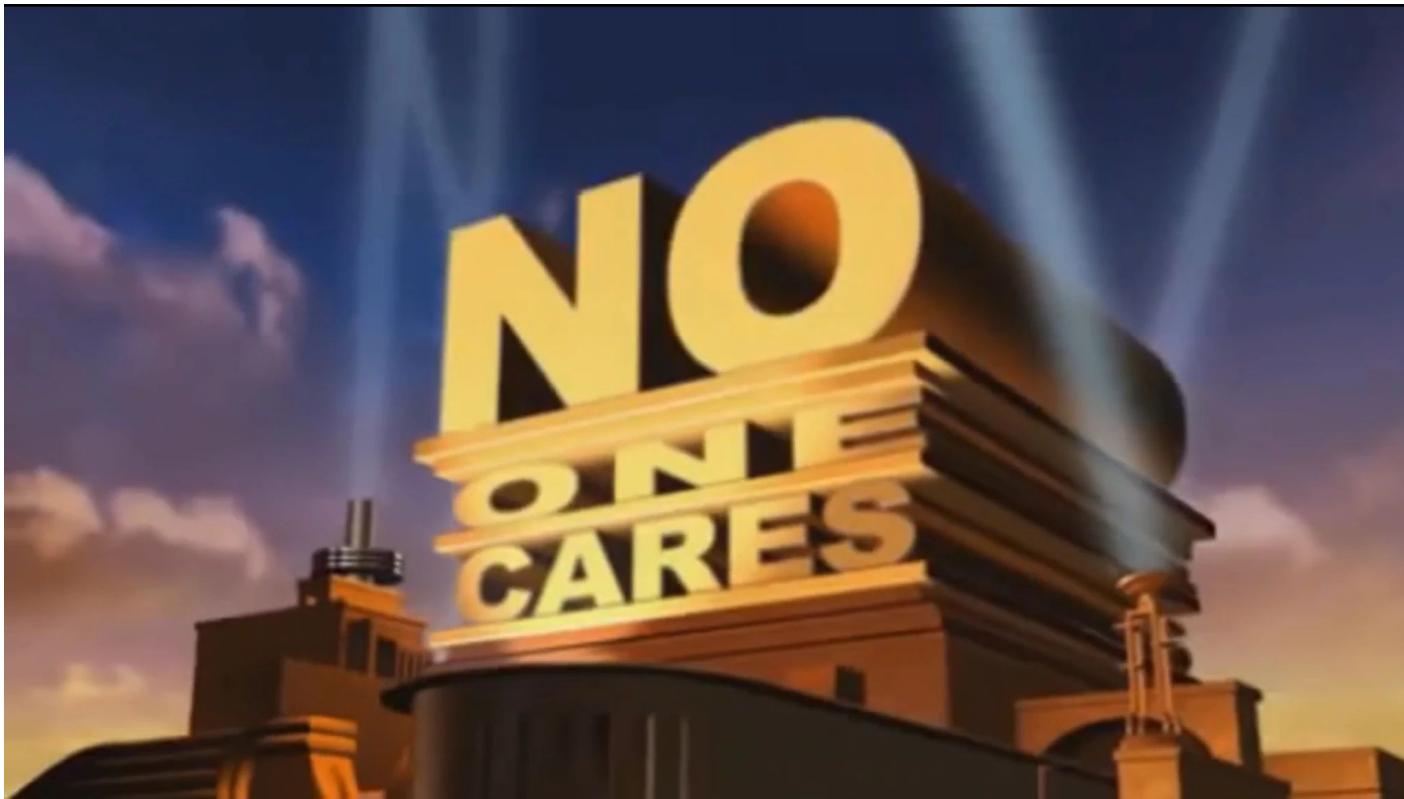
No one assigned

Labels

None yet

Projects

# November 5th, 2020: dbt-duckdb 0.18.1



# My First User ❤️



Alex Monahan

@\_\_AlexMonahan\_\_

...

@josh\_wills Hey Josh! I was just thinking about your great @duckdb @getdbt plugin. Do you think that it could also double as a parquet import/export capability? I'm thinking it could allow dbt-duckdb to be a data lake orchestrator of sorts. Maybe a hybrid w/ dbt-external-tables?

2:57 PM · Mar 14, 2021

---

2 Likes

---

# Use dbt and Duckdb instead of Spark in data pipelines



Niels Claeys · [Follow](#)

Published in datamindedbe · 7 min read · Apr 12

---

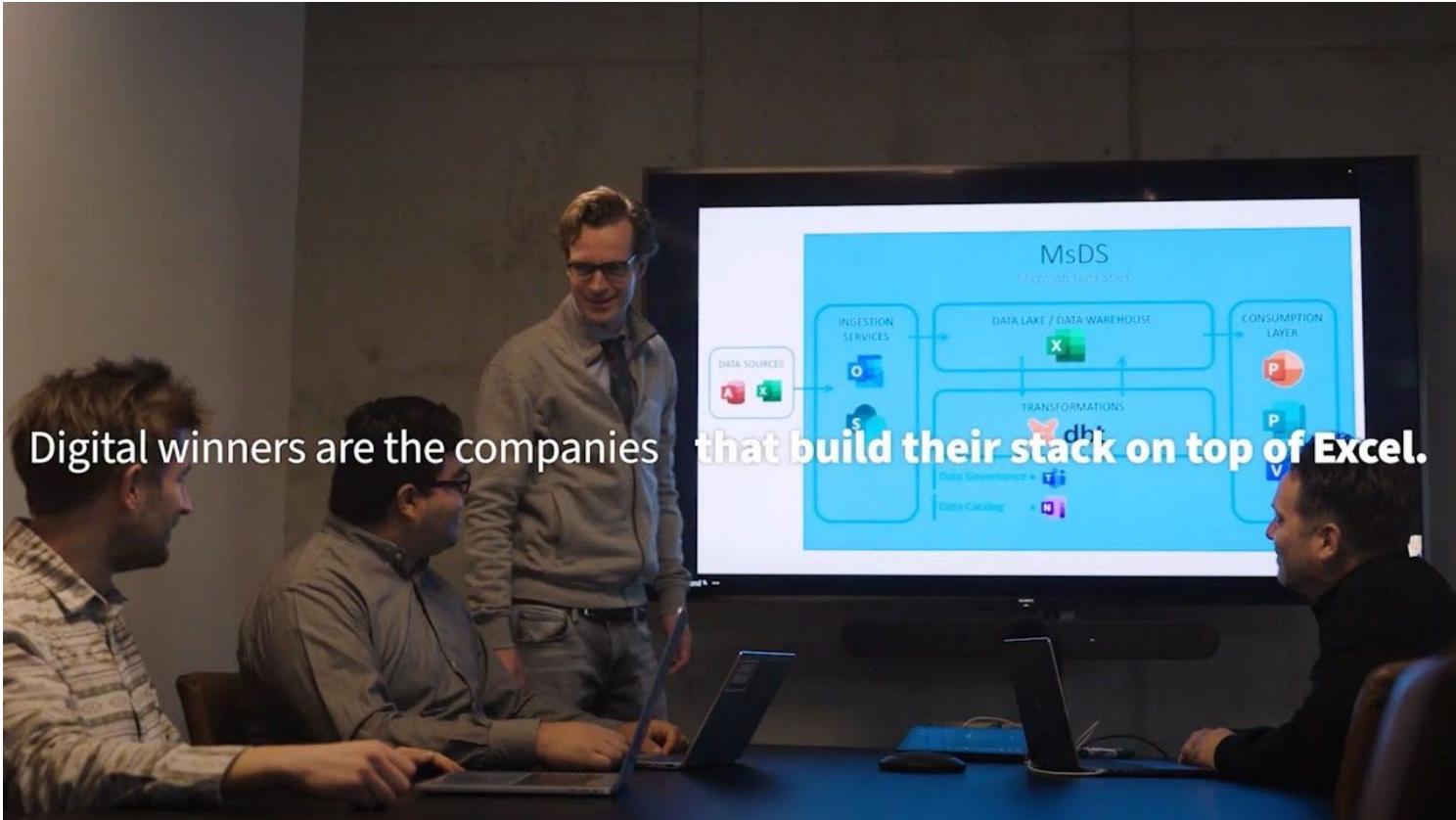
486

10



...

# dbt-excel



# April Fools'-Driven Feature Development

## dbt-duckdb should know about Excel #137

Closed

jwills opened this issue on Apr 3 · 2 comments · Fixed by #141



jwills commented on Apr 3 · edited

Owner ...

Essentially, any external file format or storage system that can read/write a `pandas.DataFrame` (or an arrow record reader, or a polars data frame, or a JSON/CSV file) and that has Python library support should be usable as either a) a source or b) an external materialization target in dbt-duckdb.

Reference: [dbt-excel.com](#)



2



4



jwills self-assigned this on Apr 3

# The Modern Data Stack in a Box

NBA Sim Dashboard

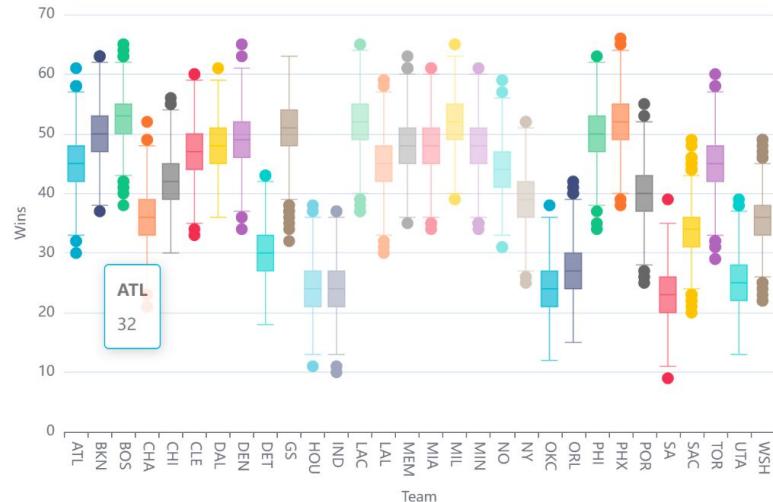
Published



### Season Summary

| conf | team | avg_wins | vegas_wins | elo_vs_vegas |
|------|------|----------|------------|--------------|
| East | DET  | 29.7     | 29.5       | -0.2         |
| East | ORL  | 26.7     | 26.5       | -0.2         |
| East | IND  | 23.9     | 23.5       | -0.4         |
| West | LAC  | 52.0     | 52.5       | 0.5          |
| West | PHX  | 51.9     | 52.5       | 0.6          |
| West | GS   | 50.9     | 51.5       | 0.6          |
| West | DEN  | 49.1     | 49.5       | 0.4          |
| West | MIN  | 48.2     | 48.5       | 0.3          |
| West | MEM  | 48.2     | 48.5       | 0.3          |
| West | DAL  | 48.1     | 48.5       | 0.4          |
| West | LAL  | 45.2     | 45.5       | 0.3          |
| West | NO   | 44.3     | 44.5       | 0.2          |
| West | POR  | 39.7     | 39.5       | -0.2         |
| West | SAC  | 33.7     | 33.5       | -0.2         |
| West | UTA  | 24.9     | 24.5       | -0.4         |
| West | HOU  | 24.0     | 23.5       | -0.5         |
| West | OKC  | 23.9     | 23.5       | -0.4         |
| West | SA   | 23.2     | 22.5       | -0.7         |

Expected Wins by Team



# Quickstart for dbt Core using GitHub Codespaces

1 Introduction

2 Related content

3 Create a codespace

4 Generate a larger data set

## Create a codespace

1. Go to the [jaffle-shop-template repository](#) after you log in to your GitHub account.
2. Click **Use this template** at the top of the page and choose **Create new repository**.
3. Click **Create repository from template** when you're done setting the options for your new repository.
4. Click **Code** (at the top of the new repository's page). Under the **Codespaces** tab, choose **Create codespace on main**. Depending on how you've configured your computer's settings, this either opens a new browser tab with the Codespace development environment with VSCode running in it or opens a new VSCode window with the codespace in it.

# Quo Vadimus?

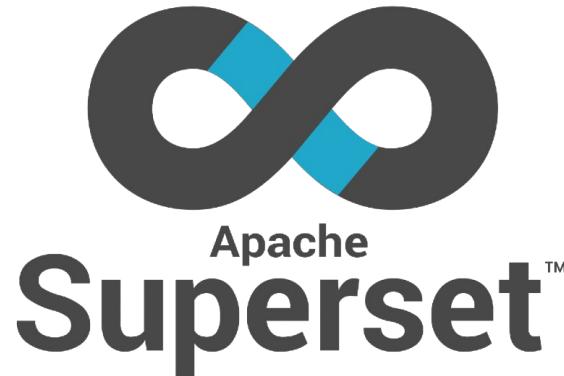
- dbt clone in 1.6.0
- Column Level Lineage via json\_serialize\_sql
- Simplifying the MDS In A Box



# Buena Vista



# DuckDBT



# trino

sQLA

# DuckDB Everywhere

Gold membership



VOLTRON DATA



MODE



MotherDuck



posit™

Silver membership



Watershed



RelationalAI



Rill

omni



LIQUID  
ANALYTICS

## SUMMER

# dbt-duckdb Grows Up/DuckDBT Stays Punk



---

**Thank you!**

---