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

- Activity Model
- Detailed Data Requirements
- Technical Environment
- Performance Considerations
- Business Data

Create/Update Logical Data Model

Conceptual Data Model
- Entities/Subtypes
- Attributes
- Relationships
- Integrity Rules

Create/Update Physical Data Model

Physical Data Model
- Tables
- Columns
- Keys/Indices
- Triggers

Create/Update Data

Data
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. 🤔 😈
My Super Villain Origin Story
In the beginning... 🇬🇧
My First Pull Request To DuckDB

Table functions `information_schema_schemata()` and `information_schema_tables()` #899

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

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 DBAver 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.
November 5th, 2020: dbt-duckdb 0.18.1
@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?
Use dbt and Duckdb instead of Spark in data pipelines
Digital winners are the companies that build their stack on top of Excel.
April Fools’-Driven Feature Development

dbt-duckdb should know about Excel #137

jwills commented on Apr 3 • edited

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

jwills self-assigned this on Apr 3
The Modern Data Stack in a Box
Quickstart for dbt Core using GitHub Codespaces

Create a codespace

1. Go to the [jaffle-shop-template](https://github.com/dbt-labs/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
DuckDB Everywhere

Gold membership

VOLTRON DATA  MODE  MotherDuck  posit

Silver membership

HEX  Watershed  RelationalAI  Rill  omni  LIQUID ANALYTICS  SUMMER
dbt-duckdb Grows Up/DuckDBT Stays Punk
Thank you!