



Introduction to the data build tool (dbt)

John Miner
Data Architect
john.miner@insight.com



Agenda

01

Background

02

Configure VS Code

03

Materializing data

04

Data Quality

05

Producing Documentation

Background

Concept

Data build tool (dbt) is an open-source tool that helps engineers **transform** existing data within their warehouses.

VS Code Integration

The framework is written in Python with the translations defined as SQL scripts and/or YAML files.

Target Platforms

The real power comes from the fact that **adapters** have been written for many different target database management systems.

Configure VS Code

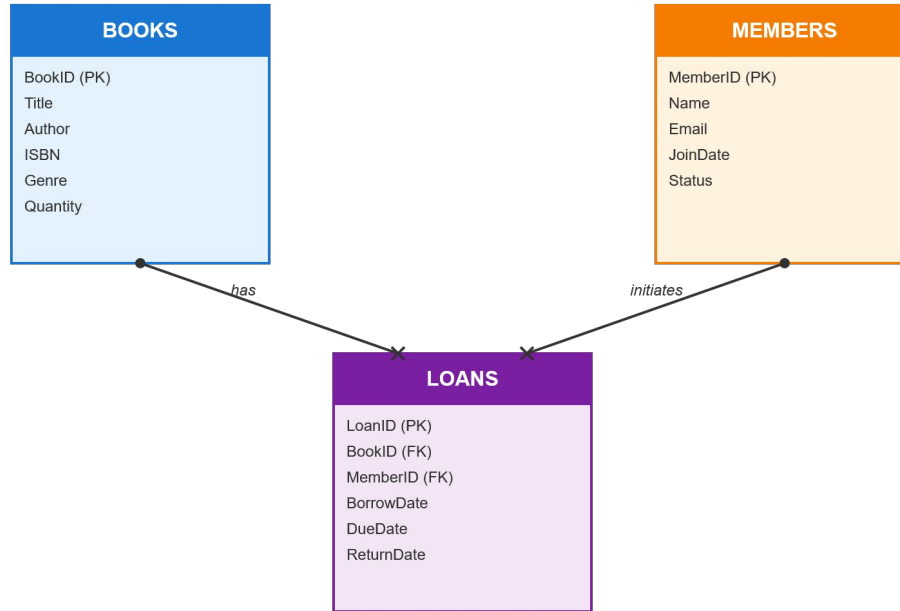
Requirements:

1. Install Python language version 3.7 or above.
2. Pip install dbt-core.
3. Pip install dbt-(adapter).
4. Local ODBC driver for (adapter).



Library Schema

- AI generated tables
- CSV sample records



Source + Target

1. Azure SQL Database
2. Initialize dbt project
3. How to seed tables



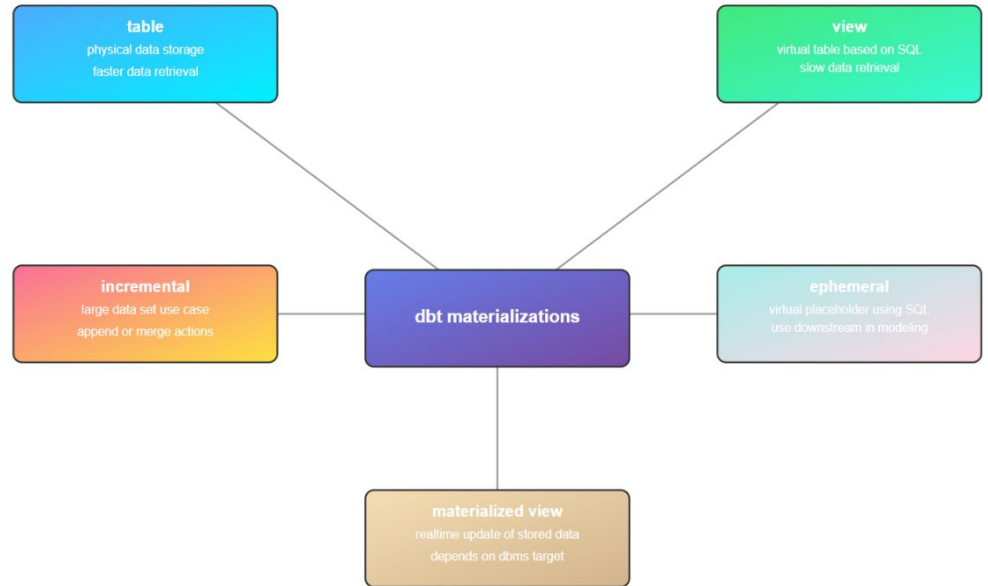
Command Line

Run commands from a terminal window.

command <input type="text"/>	argument(s) <input type="text"/>
run	<code>--select</code> , <code>--exclude</code> , <code>--selector</code> , <code>--defer</code>
test	<code>--select</code> , <code>--exclude</code> , <code>--selector</code> , <code>--defer</code>
seed	<code>--select</code> , <code>--exclude</code> , <code>--selector</code>
snapshot	<code>--select</code> , <code>--exclude</code> , <code>--selector</code>
ls (list)	<code>--select</code> , <code>--exclude</code> , <code>--selector</code> , <code>--resource-type</code>
compile	<code>--select</code> , <code>--exclude</code> , <code>--selector</code> , <code>--inline</code>
freshness	<code>--select</code> , <code>--exclude</code> , <code>--selector</code>
build	<code>--select</code> , <code>--exclude</code> , <code>--selector</code> , <code>--resource-type</code> , <code>--defer</code>
docs generate	<code>--select</code> , <code>--exclude</code> , <code>--selector</code>

Materializing Data

1. Tables
2. Views
3. Incremental
4. Ephemeral
5. Materialized Views (Not Supported)
6. Snapshots (SCD Type2)



Testing Data

1. Not Null
2. Unique
3. Accepted Values
4. Relationships
5. Custom



Generate Documentation

1. dbt docs generate
2. dbt docs serve

The screenshot displays the dbt documentation interface in a browser. The URL is localhost:8080/#/seed.library.books01. The interface shows a sidebar with a project tree where 'books01' is selected under the 'seeds' folder. The main content area displays the following information for the 'books01' seed:

- Overview:** Project and Database tabs, Group filter.
- Projects:** A tree view showing the project structure, including 'library', 'models', 'mart', 'seeds', and 'snapshots'.
- books01 seed:** Overview, Details, Description, Columns, Referenced By, SQL.
- Details:** A table with columns: TAGS (untagged), OWNER (ovacct02), TYPE (table), PACKAGE (library), RELATION (library.data_raw.books01), CONTRACT (Not Enforced).
- Description:** Raw books data.
- Columns:** A table with columns: COLUMN, TYPE, DESCRIPTION, CONSTRAINTS, DATA TESTS, MORE?.

TAGS	OWNER	TYPE	PACKAGE	RELATION	CONTRACT
untagged	ovacct02	table	library	library.data_raw.books01	Not Enforced

COLUMN	TYPE	DESCRIPTION	CONSTRAINTS	DATA TESTS	MORE?
BookID	int	Book identifier		U N	>
Title	varchar	Book title			>
Author	varchar	Book author			>
ISBN	varchar	International Standard Book Number			>
Genre	varchar	Book genre		A	>
Quantity	int	Quantity in stock			>
dates01		Raw dates data			>

Summary



The Data build tool (dbt) depends upon the data being already loaded into raw tables within the database and/or warehouse.

Power users who can write SQL and YAML can be very productive on building out models. There are several different ways to materialize the data as well as test for data quality.

The whole project is part of a directed acyclic graph. Therefore, unique file names are important. Additional information in property files allow for documentation to be produced.

In short, the dbt tool makes it easy to build out data models.

References

Chinook Sample Project

<https://tinyurl.com/yc6p7ex9>

dbt Labs – SQL Server Adapter

<https://tinyurl.com/55ez63n2>

Microsoft Fabric - Warehouse Adapter

<https://tinyurl.com/538ynn6>

Microsoft Fabric - dbt jobs

<https://tinyurl.com/yy2k8chn>