

# dbt is just SQL files, run in the right order.

D8A.ACADEMY/RESOURCES/DBT-BASICS

## THE IDEA

|                    |                  |
|--------------------|------------------|
| a model = a SELECT | one .sql file    |
| ref('stg_orders')  | never a raw name |
| dbt builds the DAG | order from refs  |

## COMMANDS

|                   |                    |
|-------------------|--------------------|
| dbt run           | build the models   |
| dbt test          | check the data     |
| dbt build         | both, in DAG order |
| dbt docs generate | lineage for free   |

## TESTS

|                  |                    |
|------------------|--------------------|
| unique, not_null | two lines of yaml  |
| accepted_values  | enum columns       |
| relationships    | FKs, checked daily |

## LAYERS

|             |                        |
|-------------|------------------------|
| staging/    | rename, cast, clean    |
| marts/      | what the BI tool reads |
| sources.yml | raw tables, declared   |

## MATERIALIZE

|             |                       |
|-------------|-----------------------|
| view        | default, always fresh |
| table       | heavy + read often    |
| incremental | big facts, append     |

## GOTCHAS

|                 |                          |
|-----------------|--------------------------|
| FROM raw.orders | ← breaks the DAG. ref(). |
| SELECT *        | schema drift upstream    |
| untested model  | silent bad joins         |

## WORTH MEMORIZING

### 01 A whole model, honestly

This file IS the table. dbt reads the refs, orders the builds, writes the DDL.

```
-- models/marts/daily_revenue.sql
SELECT
  order_date,
  SUM(amount) AS revenue,
  COUNT(DISTINCT customer_id) AS buyers
FROM {{ ref('stg_orders') }}
WHERE status = 'complete'
GROUP BY order_date
```

→ No CREATE TABLE, no DROP, no run order to remember: the ref() calls are the dependency graph.

### 02 Trap: bypassing ref()

The model works today and breaks the day the raw table moves.

| YOU WROTE                 | WHAT HAPPENS                  | WRITE INSTEAD                |
|---------------------------|-------------------------------|------------------------------|
| FROM analytics.stg_orders | invisible to the DAG          | FROM {{ ref('stg_orders') }} |
| SELECT * from staging     | upstream drift hits marts     | name the columns             |
| no tests on the mart      | bad join ships to the BI tool | unique + not_null minimum    |

→ ref() is the entire product: lineage, run order, environments. Hardcode a name and you opt out of all three.