Data Fellowship Webinars

Introduction to dbt

Nina Anderson

Session Objectives

- Learn what dbt is and why it’s useful for analysts
- Discuss 3 cool features

What is dbt?

From dbt’s creators:

dbt “enables analytics engineers to transform data in their warehouses by simply writing select statements. dbt handles turning these select statements into tables and views.”

01
02
03
04
05
06

01 Plan

02 Data prep

03 Analysis

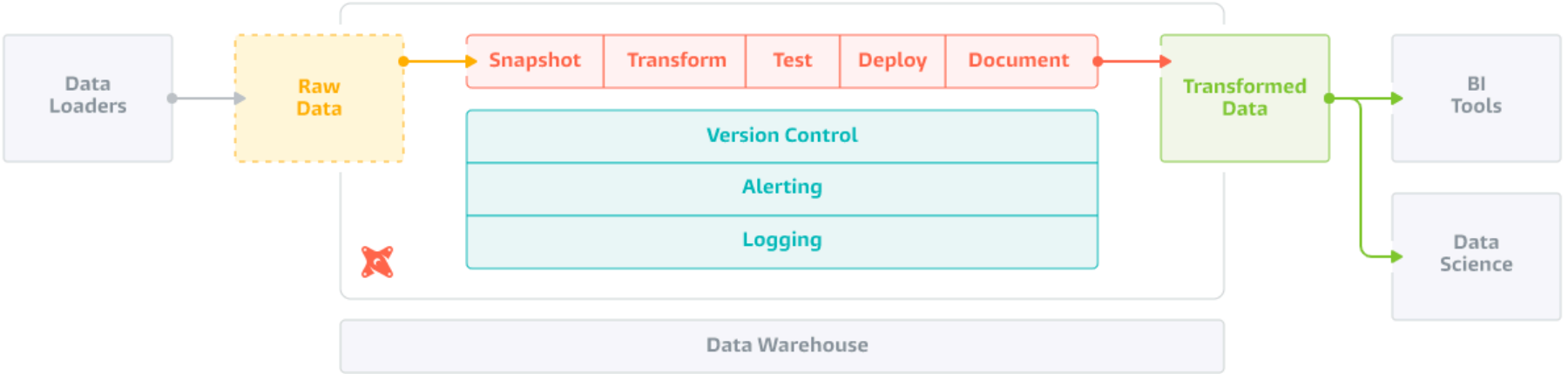
04 Modelling

05 Refine and compare

06 Communicate and implement

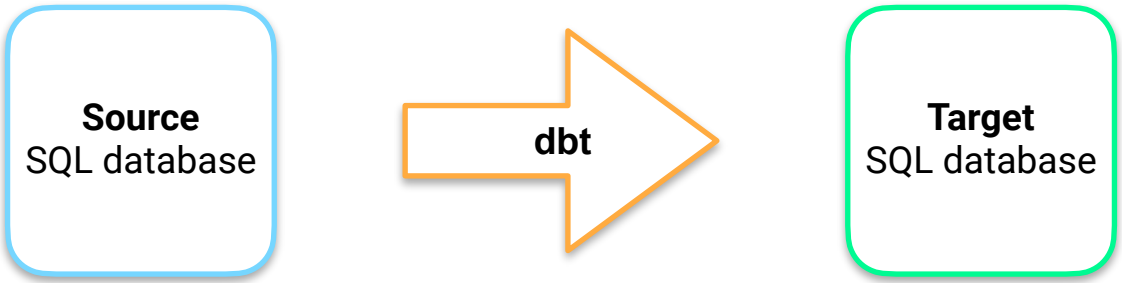
- dbt is an open source tool that you can download from the developers' website
- In addition, you need:
    - a text editor to write your code (e.g. Sublime, Atom)
    - the command line to execute that code (e.g. Terminal on Mac)

Note: there is also a paid web interface


Data Loaders
Raw Data
Snapshot
Transform
Test
Deploy
Document
Version Control
Alerting
Logging
Transformed Data
BI Tools
Data Science
Data Warehouse

How does it work?

- You write SQL select statements (plus a few extra bits)
- You write '**dbt run**' in the command line



transform

What is data transformation?

What is data transformation?

- Cleaning
- Changing structure

A real example

Cool feature #1: Incremental models

Sometimes tables have millions or billions of records

>> You might not want to transform all the old rows again

>>> We can tell dbt to **only** transform new rows using a SQL WHERE clause

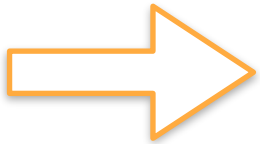
```
{% if is_incremental() %}

  -- this filter will only be applied on an incremental run
  where event_time > (select max(event_time) from {{ this }})

{% endif %}
```

Cool feature #2: Tests

Assert certain facts about the data



Check if facts are true

```yaml
models:
    - name: SESSIONS
      description: "Number of sessions and transactions per user"
      columns:
          - name: USER_ID
            tests:
              - not_null
              - unique
```

```
th dbt=0.16.0
odels, 2 tests, 0 snapshots, 0 analyses, 260 macros, 0 operations, 0 seed files, 4 sources

Concurrency: 1 threads (target='read')

1 of 2 START test not_null_SESSIONS_USER_ID..... [RUN]
1 of 2 PASS not_null_SESSIONS_USER_ID..... [PASS in 0.76s]
2 of 2 START test unique_SESSIONS_USER_ID..... [RUN]
2 of 2 FAIL 31 unique_SESSIONS_USER_ID..... [FAIL 31 in 0.79s]

Finished running 2 tests in 6.65s.

with 1 error and 0 warnings:

test unique_SESSIONS_USER_ID (models/SNOWPLOW/schema.yml)
sults, expected 0.

SQL at target/compiled/SNOWPLOW_WAREHOUSE/schema_test/unique_SESSIONS_USER_ID.sql

=1 WARN=0 ERROR=1 SKIP=0 TOTAL=2
```

Cool feature #3:
Auto Documentation

01 Plan

02 Data prep

03 Analysis

04 Modelling

05 Refine and compare

06 Communicate and implement

Resources

- Jaffle shop tutorial https://github.com/fishtown-analytics/jaffle_shop
- dbt documentation https://docs.getdbt.com/docs/building-a-dbt-project/testing-and-documentation/documentation/
- dbt community and Slack channel https://community.getdbt.com/