

I. DBT là gì?

- **DBT (Data Build Tool)** giúp **chuyển đổi dữ liệu (transformations)** trong kho dữ liệu (Data Warehouse) **theo chuẩn ELT**.
- Viết SQL dạng modular, dễ tái sử dụng.
- Quản lý version bằng Git.
- Tạo tài liệu (documentation) & lineage (sơ đồ luồng dữ liệu).

II. Flow chuẩn DBT

1. dbt init → Tạo project
2. profiles.yml → Kết nối warehouse
3. Viết models → SQL transformation
4. dbt build → Build models
5. schema.yml → Viết tests
6. dbt test → Test data quality
7. dbt docs → Generate tài liệu + lineage

III. Chi tiết sử dụng DBT

❶ Cài đặt dbt

```
pip install dbt-core dbt-postgres
```

❷ Khởi tạo project dbt

```
dbt init my_first_project  
cd my_first_project
```

❸ Cấu hình profile (kết nối database)

Tạo file ~/.dbt/profiles.yml theo mẫu:

```
my_bq_project:
```

```
target: dev
```

outputs:

dev:

type: bigquery

method: service-account

project: fsds-bi

dataset: analytics # Dataset sẽ chứa các bảng dbt tạo ra

keyfile: /Users/nguyentienhuy/Documents/fsds/BI/material/Lesson 15 - Python
3/fsds-bi-96a792b2baf8.json

threads: 4

timeout_seconds: 300

location: US # Tùy vùng của bạn

4 Tạo file model SQL

models/stg_orders.sql

SELECT

id AS order_id,

customer_id,

order_date,

total_amount

FROM

raw.orders

WHERE

status != 'cancelled'

→ Tên file là **stg_orders.sql**, dbt sẽ tạo table/views từ query này.

5 Chạy dbt build

dbt build

Khi chạy dbt build, sẽ có các step sau:

- run SQL model.
- test dữ liệu.
- generate docs.

dbt run --select customer_orders

6 Tạo tài liệu & lineage

dbt docs generate

dbt docs serve