

# IT4065C – Data Technologies Administration

**Module:** Module 2 – Data Modeling for Operational and Analytical Systems

**M2 Lab 3:** Data Modeling with dbt: Staging, Core Entities, OLTP vs OLAP

**Estimated Time:** 60–90 minutes

---

## Lab Introduction

In Module 2 Lab 2, you examined raw data and practiced data classification, focusing on:

- understanding what the data represents,
- identifying sensitive fields,
- and reasoning about governance and risk.

In Lab 3, you build on that foundation by learning how raw data is transformed into structured, trustworthy models that organizations rely on for operations and analytics.

Rather than writing SQL under pressure, this lab emphasizes:

- understanding data modeling layers,
- executing an existing pipeline correctly,
- validating outputs,
- and explaining *why* models are designed the way they are.

This mirrors how real data teams work in production environments.

---

## Learning Outcomes

By the end of this lab, you will be able to:

1. Build staging models in dbt that standardize raw data.
  2. Build core entity models (“source of truth” tables) from staging data.
  3. Explain the difference between OLTP structures (normalized) and OLAP structures (star-schema style).
  4. Use dbt tests to validate keys, constraints, and relationships.
  5. Demonstrate reasoning by justifying modeling choices and tests using governance thinking from Lab 2.
- 

## Core Principle (Read Carefully)

*All scripts required to complete this lab are already included in the repository.*

- You are **not required** to write or modify any SQL or dbt files.
- You should **not edit** any scripts unless you are intentionally exploring beyond the lab.
- Your task is to:
  - understand what exists,
  - run the correct commands,
  - verify results,
  - and reason about the modeling decisions.

If something does not work as expected, **do not guess or change files**. Compare your output to the expected outcomes and notify the instructor or TA if needed.

---

## Section 1: Required Files Checklist (Verify – Do Not Edit)

### Instructions to students:

Before running any commands, ensure that the following files are present in your cloned repository.

If *any file is missing*, stop and notify the instructor or TA.

#### 1. Lab Instruction & Validation Files

File	Purpose
labs/module_2/lab3/README.md	Official lab instructions
labs/module_2/lab3/lab3_quick_checks.sql	Prewritten SQL queries to verify dbt outputs
labs/module_2/lab3/lab3_turnin_template.md	Submission structure and reflection prompts

## 2. Raw Data Dependency (Provided Ahead of Time)

File	Purpose
labs/module_2/lab2_seed.sql	Loads raw tables required for modeling

### Why this exists:

Lab 3 must be runnable even if Lab 2 has not yet been completed.

## 3. dbt Modeling Files (Read-Only for This Lab)

Staging models clean, standardize, and lightly transform raw source data. They provide consistent column names and formats for downstream models and should not contain business logic.

### Staging Models (Standardization Layer)

File	Purpose
IT4065C-Labs/dbt/it4065c_platform/models/staging/lab3/stg_customers.sql	Cleans and standardizes raw customer data
IT4065C-Labs/dbt/it4065c_platform/models/staging/lab3/stg_orders.sql	Standardizes order records (order headers)
IT4065C-Labs/dbt/it4065c_platform/models/staging/lab3/stg_products.sql	Standardizes product information
IT4065C-Labs/dbt/it4065c_platform/models/staging/lab3/stg_order_items.sql	Standardizes line-item level transaction data

## 4. Core Models (Trusted Entities)

Core models represent trusted business entities and facts. These tables enforce primary keys, relationships, and are the recommended layer for most analysis.

File	Purpose
IT4065C-Labs/dbt/it4065c_platform/models/core/lab3/dim_customers.sql	Customer dimension (one row per customer)
IT4065C-Labs/dbt/it4065c_platform/models/core/lab3/dim_products.sql	Product dimension
IT4065C-Labs/dbt/it4065c_platform/models/core/lab3/fct_orders.sql	Orders fact table (one row per order)
IT4065C-Labs/dbt/it4065c_platform/models/core/lab3/fct_order_items.sql	Line-item fact table (one row per order item)

## 5. Analytical Views (Usage-Oriented Models)

Mart models are usage-oriented and optimized for specific query patterns. They are built on top of the core models and demonstrate different analytical styles.

File	Purpose
IT4065C-Labs/dbt/it4065c_platform/models/marts/lab3/oltp_order_detail.sql	OLTP-style normalized view for detailed order inspection
IT4065C-Labs/dbt/it4065c_platform/models/marts/lab3/olap_sales_by_day.sql	OLAP-style aggregated view for daily sales analysis

## 6. Tests & Governance Rules

This file defines dbt schema tests and documentation for all Lab 3 models. Students do not need to edit this file.

File	Purpose
IT4065C-Labs/dbt/it4065c_platform/models/lab3/_lab3_models.yml	dbt tests for primary keys, uniqueness, and foreign-key relationships

### Key idea:

Data integrity is enforced in dbt without modifying the database schema.

## 7. Optional Automation (If Included)

File	Purpose
labs/module_2/lab3/lab3_run_all.sh	Runs dbt + validation in one command

### Student Action

✓ Confirm all files exist

✗ Do not edit any files

📸 **Screenshot 1 (Required):** directory listing showing Lab 3 files exist.

## Section 2: What You Will Run (Commands Only)

### Step 2.1: Activate Environment

```
cd ~/IT4065C-Labs  
source dbt-venv/bin/activate  
cd dbt/it4065c_platform
```

```
kofi@18ntiikxil02:~/IT4065C- ~ | +  
kofi@18ntiikxil02:~/IT4065C-Labs$ cd ~/IT4065C-Labs  
kofi@18ntiikxil02:~/IT4065C-Labs$ source dbt-venv/bin/activate  
(dbt-venv) kofi@18ntiikxil02:~/IT4065C-Labs$ cd dbt/it4065c_platform  
(dbt-venv) kofi@18ntiikxil02:~/IT4065C-Labs/dbt/it4065c_platform$
```

### Step 2.2: Validate dbt Configuration

dbt debug

Expected outcome: database connection test passes.

⌚ Screenshot 2: console output.

```
kofi@18ntiikxil02:~/IT4065C- ~ | +  
(dbt-venv) kofi@18ntiikxil02:~/IT4065C-Labs/dbt/it4065c_platform$ dbt debug  
03:34:02 Running with dbt=1.11.2  
03:34:02 ddt version: 1.11.2  
03:34:02 python version: 3.10.12  
03:34:02 python path: /home/kofi/IT4065C-Labs/dbt-venv/bin/python3  
03:34:02 os info: Linux-5.15.107.4-microsoft-standard-WSL2-x86_64-with-glibc2.35  
03:34:02 dbt project root: /home/kofi/IT4065C-Labs/dbt/it4065c_platform  
03:34:02 Using profiles.yml file at /home/kofi/IT4065C-Labs/dbt/it4065c_platform/profiles.yml  
03:34:02 Using dbt_project.yml file at /home/kofi/IT4065C-Labs/dbt/it4065c_platform/dbt_project.yml  
03:34:02 adapter type: postgres  
03:34:02 adapter version: 1.10.0  
03:34:02 Configuration files found:  
03:34:02   - dbt_profiles.yml file [OK Found and valid]  
03:34:02   - dbt_project.yml file [OK Found and valid]  
03:34:02 Required dependencies:  
03:34:02   - git [OK Found]  
  
03:34:02 Connection:  
03:34:02   host: localhost  
03:34:02   port: 5432  
03:34:02   database: it4065c_platform  
03:34:02   schema: student_kofi  
03:34:02   connect_timeout: 10  
03:34:02   role: None  
03:34:02   search_path: None  
03:34:02   host_encoding: idle  
03:34:02   sslmode: prefer  
03:34:02   sslcert: None  
03:34:02   sslkey: None  
03:34:02   sslcrl: None  
03:34:02   application_name: dbt  
03:34:02   retries: 1  
03:34:02 Registered adapter: postgres=1.10.0  
03:34:02 Connection test: [OK connection ok]  
03:34:02 All checks passed!  
(dbt-venv) kofi@18ntiikxil02:~/IT4065C-Labs/dbt/it4065c_platform$
```

### Step 2.3: Build Lab 3 Models Only

dbt run --select path:models/lab3

Expected outcome:

Staging → Core → Marts models build successfully.

⌚ Screenshot 3: console output.

```
kofi@18ntiikxil02:~/IT4065C- ~ | +  
(dbt-venv) kofi@18ntiikxil02:~/IT4065C-Labs/dbt/it4065c_platform$ dbt run --select path:models/lab3  
04:52:31 Running with dbt=1.11.2  
04:52:31 Registered adapter: postgres=1.10.0  
04:52:31 [WARNING] [MissingArgumentsPropertyInGenericTestDeprecation]: Deprecated functionality.  
Found top-level arguments to test 'relationships' defined on 'stg_order_items'  
in package 'it4065c_platform' ('models/lab3/_lab3_models.yml'). Arguments to  
generic tests must be nested under the 'arguments' property.  
04:52:32 [WARNING] [MissingArgumentsPropertyInGenericTestDeprecation]: There are 3 unused configurations in your dbt_project.yml file which do not apply to any resources.  
There are 3 unused configuration paths:  
- models.it4065c_platform.marts  
- models.it4065c_platform.core  
- models.it4065c_platform.streaming  
04:52:32 Found 10 models, 28 data tests, 464 macros  
04:52:32 Concurrency: 2 threads (target='dev')  
04:52:32 1 of 10 OK START sql view model student_kofi.stg_customers ..... [RUN]  
04:52:32 2 of 10 OK START sql view model student_kofi.stg_order_items ..... [RUN]  
04:52:32 2 of 10 OK created sql view model student_kofi.stg_order_items ..... [CREATE VIEW in 0.17s]  
04:52:32 3 of 10 OK created sql view model student_kofi.stg_order_items ..... [RUN]  
04:52:32 4 of 10 OK created sql view model student_kofi.stg_products ..... [CREATE VIEW in 0.19s]  
04:52:32 3 of 10 OK created sql view model student_kofi.stg_products ..... [RUN]  
04:52:32 4 of 10 OK created sql view model student_kofi.stg_products ..... [CREATE VIEW in 0.05s]  
04:52:32 6 of 10 OK created sql view model student_kofi.dim_customers ..... [CREATE VIEW in 0.05s]  
04:52:32 6 of 10 OK created sql view model student_kofi.dim_customers ..... [RUN]  
04:52:32 7 of 10 OK created sql view model student_kofi.dim_products ..... [CREATE VIEW in 0.11s]  
04:52:32 8 of 10 OK created sql view model student_kofi.dim_products ..... [RUN]  
04:52:32 8 of 10 OK created sql view model student_kofi.dim_products ..... [CREATE VIEW in 0.06s]  
04:52:32 9 of 10 OK created sql view model student_kofi.dim_products ..... [RUN]  
04:52:32 9 of 10 OK created sql view model student_kofi.olap_sales_by_day ..... [CREATE VIEW in 0.05s]  
04:52:32 10 of 10 OK START sql view model student_kofi.oltp_order_detail ..... [RUN]  
04:52:33 10 of 10 OK created sql view model student_kofi.oltp_order_detail ..... [CREATE VIEW in 0.10s]  
04:52:33 9 of 10 OK created sql view model student_kofi.olap_sales_by_day ..... [CREATE VIEW in 0.10s]  
04:52:33 Finished running 10 view models in 0 hours 0 minutes and 0.74 seconds (0.74s).  
04:52:33 Completed successfully  
04:52:33 [RUN]  
04:52:33 Done. PASS=10 WARN=0 ERROR=0 SKIP=0 NO_OP=0 TOTAL=10  
04:52:33 [WARNING] [DeprecationsSummary]: Deprecated functionality.  
Summary of deprecated functionality:  
- MissingArgumentsPropertyInGenericTestDeprecation: 2 occurrences  
To see all depreciation instances instead of just the first occurrence of each,  
run command again with the --show-all-deprecations flag. You may also need to  
use the --no-partial-parse as some deprecations are only encountered during  
parsing.  
(dbt-venv) kofi@18ntiikxil02:~/IT4065C-Labs/dbt/it4065c_platform$
```

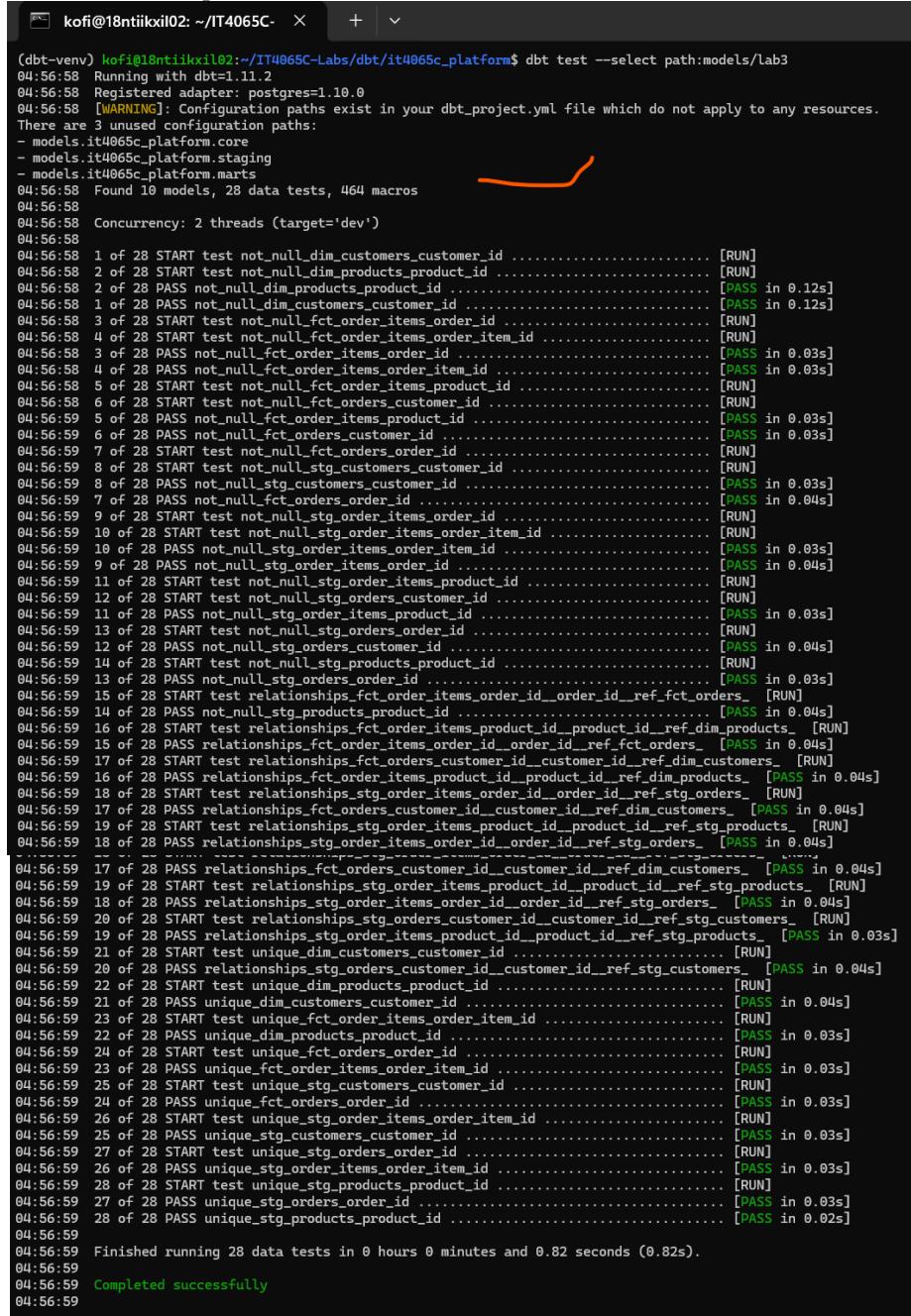
## Step 2.4: Run Data Quality Tests

dbt test --select path:models/lab3

### Expected outcome:

- All tests pass
- No orphan foreign keys
- No duplicate primary keys

 Screenshot 4: console output.



```
(dbt:venv) kofi@18ntiikxi02:~/IT4065C-Labs/dbt/it4065c_platform$ dbt test --select path:models/lab3
04:56:58 Running with dbt=1.11.2
04:56:58 Registered adapter: postgres=1.10.0
04:56:58 [WARNING]: Configuration paths exist in your dbt_project.yml file which do not apply to any resources.
There are 3 unused configuration paths:
- models.it4065c.platform.core
- models.it4065c.platform.staging
- models.it4065c.platform.marts
04:56:58 Found 10 models, 28 data tests, 464 macros
04:56:58 Concurrency: 2 threads (target='dev')
04:56:58
04:56:58 1 of 28 START test not_null_dim_customers_customer_id ..... [RUN]
04:56:58 2 of 28 START test not_null_dim_products_product_id ..... [RUN]
04:56:58 2 of 28 PASS not_null_dim_products_product_id ..... [PASS in 0.12s]
04:56:58 1 of 28 PASS not_null_dim_customers_customer_id ..... [PASS in 0.12s]
04:56:58 3 of 28 START test not_null_fct_order_items_order_id ..... [RUN]
04:56:58 4 of 28 START test not_null_fct_order_items_order_item_id ..... [RUN]
04:56:58 3 of 28 PASS not_null_fct_order_items_order_id ..... [PASS in 0.03s]
04:56:58 4 of 28 PASS not_null_fct_order_items_order_item_id ..... [PASS in 0.03s]
04:56:58 5 of 28 START test not_null_fct_order_items_product_id ..... [RUN]
04:56:58 6 of 28 START test not_null_fct_orders_customer_id ..... [RUN]
04:56:59 5 of 28 PASS not_null_fct_order_items_product_id ..... [PASS in 0.03s]
04:56:59 6 of 28 PASS not_null_fct_orders_customer_id ..... [PASS in 0.03s]
04:56:59 7 of 28 START test not_null_stg_orders_order_id ..... [RUN]
04:56:59 8 of 28 START test not_null_stg_customers_customer_id ..... [RUN]
04:56:59 8 of 28 PASS not_null_stg_customers_customer_id ..... [PASS in 0.03s]
04:56:59 7 of 28 PASS not_null_stg_orders_order_id ..... [PASS in 0.04s]
04:56:59 9 of 28 START test not_null_stg_order_items_order_id ..... [RUN]
04:56:59 10 of 28 START test not_null_stg_order_items_order_item_id ..... [PASS in 0.03s]
04:56:59 9 of 28 PASS not_null_stg_order_items_order_id ..... [PASS in 0.04s]
04:56:59 11 of 28 START test not_null_stg_order_items_product_id ..... [RUN]
04:56:59 12 of 28 START test not_null_stg_orders_customer_id ..... [RUN]
04:56:59 11 of 28 PASS not_null_stg_order_items_product_id ..... [PASS in 0.03s]
04:56:59 13 of 28 START test not_null_stg_orders_order_id ..... [RUN]
04:56:59 12 of 28 PASS not_null_stg_orders_customer_id ..... [PASS in 0.04s]
04:56:59 14 of 28 START test not_null_stg_products_product_id ..... [RUN]
04:56:59 13 of 28 PASS not_null_stg_orders_order_id ..... [PASS in 0.03s]
04:56:59 15 of 28 START test relationships_fct_order_items_order_id_ref_fct_orders_ ..... [RUN]
04:56:59 14 of 28 PASS not_null_stg_products_product_id ..... [PASS in 0.04s]
04:56:59 16 of 28 START test relationships_fct_order_items_product_id_ref_dim_products_ ..... [RUN]
04:56:59 15 of 28 PASS relationships_fct_order_items_order_id_order_id_ref_fct_orders_ ..... [PASS in 0.04s]
04:56:59 17 of 28 START test relationships_fct_orders_customer_id_customer_id_ref_dim_customers_ ..... [RUN]
04:56:59 16 of 28 PASS relationships_fct_order_items_product_id_product_id_ref_dim_products_ ..... [PASS in 0.04s]
04:56:59 18 of 28 START test relationships_stg_order_items_order_id_order_id_ref_stg_orders_ ..... [RUN]
04:56:59 17 of 28 PASS relationships_fct_orders_customer_id_customer_id_ref_dim_customers_ ..... [PASS in 0.04s]
04:56:59 19 of 28 START test relationships_stg_order_items_product_id_product_id_ref_stg_products_ ..... [RUN]
04:56:59 18 of 28 PASS relationships_stg_order_items_order_id_order_id_ref_stg_orders_ ..... [PASS in 0.04s]
04:56:59 17 of 28 PASS relationships_stg_order_items_order_id_order_id_ref_stg_orders_ ..... [PASS in 0.04s]
04:56:59 20 of 28 START test relationships_stg_orders_order_id_order_id_ref_stg_customers_ ..... [RUN]
04:56:59 19 of 28 PASS relationships_stg_order_items_product_id_product_id_ref_stg_products_ ..... [PASS in 0.03s]
04:56:59 21 of 28 START test unique_dim_customers_customer_id ..... [RUN]
04:56:59 20 of 28 PASS relationships_stg_orders_customer_id_customer_id_ref_stg_customers_ ..... [PASS in 0.04s]
04:56:59 22 of 28 START test unique_dim_products_product_id ..... [RUN]
04:56:59 21 of 28 PASS unique_dim_customers_customer_id ..... [PASS in 0.04s]
04:56:59 23 of 28 START test unique_fct_order_items_order_item_id ..... [RUN]
04:56:59 22 of 28 PASS unique_dim_products_product_id ..... [PASS in 0.03s]
04:56:59 24 of 28 START test unique_fct_orders_order_id ..... [RUN]
04:56:59 23 of 28 PASS unique_fct_order_items_order_item_id ..... [PASS in 0.03s]
04:56:59 25 of 28 START test unique_stg_order_items_order_item_id ..... [RUN]
04:56:59 24 of 28 PASS unique_fct_orders_order_id ..... [PASS in 0.03s]
04:56:59 26 of 28 START test unique_stg_order_items_order_item_id ..... [RUN]
04:56:59 25 of 28 PASS unique_stg_customers_customer_id ..... [PASS in 0.03s]
04:56:59 27 of 28 START test unique_stg_orders_order_id ..... [RUN]
04:56:59 26 of 28 PASS unique_stg_order_items_order_item_id ..... [PASS in 0.03s]
04:56:59 28 of 28 START test unique_stg_products_product_id ..... [RUN]
04:56:59 27 of 28 PASS unique_stg_orders_order_id ..... [PASS in 0.03s]
04:56:59 28 of 28 PASS unique_stg_products_product_id ..... [PASS in 0.02s]
04:56:59
04:56:59 Finished running 28 data tests in 0 hours 0 minutes and 0.82 seconds (0.82s).
04:56:59
04:56:59 Completed successfully
04:56:59
```

## Step 2.5: Validate Outputs (No SQL Writing)

cd ~/IT4065C-Labs

psql -h localhost -U postgres -d it4065c -f labs/module\_2/lab3/lab3\_quick\_checks.sql

### Expected outcome:

- Reasonable (non-zero) row counts
- OLAP view returns aggregated results
- OLTP view returns detailed rows

Exact numbers may differ slightly. You are checking structure and existence, not exact totals.

### Screenshot 5: console output.

order_id	order_date	order_status	total_amount	payment_method	customer_id	first_name	last_name	email	pho
ne_number	order_item_id	product_id	quantity	unit_price	line_total	product_name	category	price	
-555-1023	1	Completed	1	29.99	79.98	Credit Card	Electronics	1   Alice	Johnson   alice.johnson@email.com   513
-555-1023	1	Completed	1	29.99	79.98	Wireless Mouse	Electronics	1   Alice	Johnson   alice.johnson@email.com   513
-555-1023	2	Completed	1	49.99	49.99	Laptop Stand	Electronics	1   Alice	Johnson   alice.johnson@email.com   513
-555-1023	2	Completed	1	49.99	49.99	Gift Card	Electronics	2   Brian	Smith   brian.smith@email.com   513
-555-1023	3	Completed	1	49.99	49.99	Laptop Stand	Electronics	2   Brian	Smith   brian.smith@email.com   513
-555-7781	3	Cancelled	1	19.95	19.95	Store Credit	Home Goods	3   Carla	Nguyen   carla.nguyen@email.com   513
-555-1023	4	Completed	1	19.95	19.95	Water Bottle	Home Goods	3   Carla	Nguyen   carla.nguyen@email.com   513
-555-1023	4	Completed	1	34.99	34.99	Credit Card	Office Supplies	1   Alice	Johnson   alice.johnson@email.com   513
-555-1023	5	Completed	2	9.98	19.96	Notebook	Office Supplies	1   Alice	Johnson   alice.johnson@email.com   513

- If the same order\_id appears on multiple rows, that is expected, this view is one row per order item.
- Governance Note: email and phone\_number are PII. In real organizations, access would be restricted (RBAC), audited, and monitored.
- price is the product's standard price; unit\_price is the price actually paid at purchase time.
- **Note:** If you see (END) at the bottom of the screen, press q to return to the prompt.

---

## Section 3: Critical Thinking & Reasoning

Answer the following questions in 3–7 sentences per question.

### Q1: Modeling Intent

- Why does this lab separate staging models from core models instead of modeling directly from raw tables?

### Q2: OLTP vs OLAP

- Compare oltp\_order\_detail and olap\_sales\_by\_day.
- Which would you use for a dashboard?
- Which for transaction troubleshooting?
- Why?

### Q3: Data Integrity Thinking

- Pick one dbt test from \_lab3\_models.yml.
- What real-world data issue would this test help detect or prevent?

### Q4: Governance Continuity

- Based on what you learned in Lab 2, which field in this lab would require restricted access or monitoring in a real organization?
- Where would that control be enforced?

### Q5 (Optional): Advanced Thought

- If this dataset grew to billions of records, which part of the model design would need to change first, and why?

## Deliverables

Submit **one PDF or Word document** containing:

1. Screenshot 1 – Required files exist
2. Screenshot 2 – dbt debug
3. Screenshot 3 – dbt run
4. Screenshot 4 – dbt test
5. Screenshot 5 – Validation query results
6. Written responses to all reflection questions

This lab is about understanding, validating, and reasoning about data models, not writing code under pressure.

Focus on:

- Structure
- Integrity
- Purpose
- Governance

That is how real data systems are evaluated.