

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



HW3 - DATA GENERATION

MÔN	HỌC:	KIỂM	THỬ	PHẦN	MỀM
------------	-------------	-------------	------------	-------------	------------

SINH VIÊN:

Nguyễn Phúc An - 22127002

GIẢNG VIÊN:

Thầy Lâm Quang Vũ

Thầy Hồ Tuấn Thanh

Thầy Trương Phước Lộc.

Thứ 2, ngày 21 tháng 7 năm 2025

Outline

Heading 1.....	2
1. Data generation.....	3
1.1 Data fields and randomized ranges / rules.....	3
1.2 Screenshots & Process.....	9

1. Data generation

1.1 Data fields and randomized ranges / rules

Users

Field	Ranges / Rules
Id	Tăng dần , từ 1 - 100
First Name	Họ phổ biến
Last Name	Tên phổ biến
Address	Địa chỉ gồm tên và số
City	Các thành phố nổi tiếng trên thế giới
State	Áp dụng nếu quốc gia đó có hệ thống bang, ngược lại thì NULL
Country	Khớp với City
Postal code	Ứng với mã code của từng quốc gia
Phone	Là chuỗi số gồm 11-12 ký tự
DOB	Ngày sinh trong quá khứ (trước ngày hiện tại)
Email	Theo định dạng abc@example.com
Password	Chuỗi được mã hóa
Role	user
Enabled	1
Failed_login_attempt	0
Created_at	NULL
Updated_at	NULL

Products

Field	Ranges / Rules
Id	Tăng dần, từ 1 - 300
Name	Tên sản phẩm, mô tả ngắn gọn
Description	Mô tả chi tiết về sản phẩm
Stock	Số nguyên, từ 0 trở lên
Price	Số thực, định dạng tiền tệ
is_location_offer	Boolean (0 hoặc 1), 1 nếu sản phẩm có ưu đãi theo địa điểm, ngược lại là 0
is_rental	Boolean (0 hoặc 1), 1 nếu sản phẩm có thể cho thuê, ngược lại là 0
brand_id	Số nguyên, tham chiếu đến Id của bảng Brand
category_id	Số nguyên, tham chiếu đến Id của bảng Category
product_image_id	Số nguyên, tham chiếu đến Id của bảng Product Image
Created_at	Ngày giờ trong quá khứ (trước ngày hiện tại), theo định dạng YYYY-MM-DD HH:MM:SS
Updated_at	Ngày giờ trong quá khứ hoặc hiện tại, sau Created_at, theo định dạng YYYY-MM-DD HH:MM:SS

Product_image

Field	Ranges / Rules
id	Tăng dần, từ 1 - 500 \
url	Chuỗi URL hợp lệ đến hình ảnh sản phẩm
by_url	Boolean (0 hoặc 1)
source_name	Unplash
source_url	Chuỗi URL hợp lệ đến trang nguồn của hình ảnh
file_name	Tên tệp của hình ảnh theo định dạng jpg
title	Tiêu đề hoặc mô tả ngắn gọn về hình ảnh
created_at	Ngày giờ trong quá khứ (trước ngày hiện tại) theo định dạng YYYY-MM-DD HH:MM:SS
updated_at	Ngày giờ trong quá khứ hoặc hiện tại, sau Created_at, theo định dạng YYYY-MM-DD HH:MM:SS

Category

Field	Rules/Ranges
id	Tăng dần, 1-10
parent_id	ID Category cha hoặc NULL
name	Phổ biến
slug	Theo name. Nếu name gồm nhiều từ thì viết thường cách nhau dấu -
created_at	Ngày giờ quá khứ
updated_at	Ngày giờ sau Created_at

Invoice items

Field	Rules/Ranges
id	Tăng dần, 1 - 100
invoice_id	Tăng dần, tương ứng với id (mỗi hóa đơn chỉ có 1 sản phẩm)
product_id	ID Product duy nhất, tham chiếu Product
unit_price	Số thực, giá của sản phẩm tại thời điểm bán
quantity	Luôn là 1
created_at	Ngày giờ quá khứ
updated_at	Ngày giờ sau Created_at hoặc NULL

Invoice

Field	Rules/Ranges
id	Tăng dần, 1 - 100
user_id	ID User hợp lệ (tham chiếu bảng Users)
invoice_date	Ngày giờ trong quá khứ hoặc hiện tại
invoice_number	Duy nhất, random từ 1-100
billing_address	Địa chỉ thanh toán (tương tự địa chỉ User)
billing_city	Thành phố thanh toán (tương tự City của User). Nếu không có thì NULL
billing_state	Bang/Tỉnh thanh toán (tương tự State của User) Nếu không có thì NULL
billing_country	Quốc gia thanh toán (tương tự Country của User)
billing_postcode	Mã bưu chính thanh toán (tương tự Postal code của User)
total	Số thực, bằng unit_price của sản phẩm trong Invoice Item tương ứng
payment_method	Gồm "Cash on Delivery", "Bank transfer"
payment_account_name	Tên tài khoản thanh toán (nếu có)
payment_account_number	Số tài khoản thanh toán (nếu có)
status	"AWAITING_FULFILLMENT",
status_message	Tin nhắn trạng thái bổ sung hoặc NULL
created_at	Ngày giờ quá khứ
updated_at	Ngày giờ sau Created_at hoặc NULL

Brands

Field	Rules/Ranges
id	Tăng dần, 1 - 200 (hoặc số lượng thương hiệu dự kiến)
name	Tên thương hiệu (ví dụ: "NovaTech Solutions", "GreenLeaf Organics")
slug	Chuỗi duy nhất, phiên bản thân thiện với URL của tên thương hiệu (ví dụ: "novatech-solutions"). Thường là tên viết thường, không dấu, dấu cách thay bằng dấu gạch ngang.
created_at	Ngày giờ trong quá khứ hoặc hiện tại (trước hoặc bằng ngày hiện tại)
updated_at	Ngày giờ sau Created_at hoặc bằng Created_at

Contact replies

Field	Rules/Ranges
id	Tăng dần, từ 1 - 100
message_id	Số nguyên, tham chiếu đến ID bảng message
user_id	Số nguyên, tham chiếu đến ID của người dùng đã phản hồi (từ bảng Users), có thể lặp lại
message	Chuỗi văn bản có ý nghĩa, nội dung phản hồi (ví dụ: "Cảm ơn bạn đã liên hệ!", "Chúng tôi đã xử lý yêu cầu của bạn.")
created_at	Ngày giờ trong quá khứ hoặc hiện tại (trước hoặc bằng thời gian hiện tại)

updated_at	Ngày giờ sau Created_at hoặc bằng Created_at, hoặc NULL (nếu chưa được cập nhật)
------------	--

Công cụ sử dụng:

1. Gemini:

- Tạo các bảng: `categories`, `products`, `product_images`, `invoices`, `invoice_items`, và `contact_request_replies`.
- Giúp sinh ra các tên phù hợp với ngữ cảnh toolshop (categories, brands), và liên kết các `products` với `product_images` hợp lý, `invoice_items` với `invoices`, `products` theo liên kết khóa ngoại đúng logic.
- Cho phép xuất dữ liệu ra file Google Sheets để dễ dàng quản lý và chia sẻ.

2. Python (sử dụng thư viện Faker):

- Tạo bảng `brands` (100 dòng) và `users` (100 dòng).
- Giúp giữ đúng logic ngữ cảnh trong toolshop, vừa sử dụng khác công cụ với các thành viên trong nhóm, đồng thời số lượng lớn các dòng dữ liệu này được sinh ra hợp lý và linh hoạt nhờ thư viện Faker.

1.2 Screenshots & Process

Công cụ sử dụng : Gemini

Sử dụng docker để truy cập vào trang PHP Admin và truy xuất các bảng dữ liệu tương ứng. Khi truy cập được các trường dữ liệu cần thiết của từng bảng, khi này ta sẽ screenshots lại bảng dữ liệu và prompt cho Gemini sinh dữ liệu cho từng bảng theo ảnh với số lượng theo yêu cầu đề bài .

Extra options

	id	name	slug	created_at	updated_at
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	1	Brand name 1	brand-name-1	NULL	NULL
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	2	Brand name 2	brand-name-2	NULL	NULL

tao du lieu fake 100 brands theo hinh

- ▶ Certainly! Here's the data separated by commas (CSV format) so you can easily paste it into Google Sheets. Each line represents a row, and commas separate the columns.


Code snippet



```
id,name,slug,created_at,updated_at
1,Giải pháp TechNova,giai-phap-technova,2025-07-22 10:25:20,2025-07-22 10:25:20
2,Đổi mới Sinh thái,doi-moi-sinh-thai,2025-07-22 10:25:20,2025-07-22 10:25:20
3,Thiết bị Hiệu suất Đỉnh cao,thiet-bi-hieu-suat-dinh-cao,2025-07-22 10:25:20,2025-07-22 10:25:20
4,Sản phẩm Thủ công Mỹ nghệ,san-pham-thu-cong-my-nghe,2025-07-22 10:25:20,2025-07-22 10:25:20
5,Viễn thông Kết nối Toàn cầu,vien-thong-ket-noi-toan-cau,2025-07-22 10:25:20,2025-07-22 10:25:20
6,Thời trang Tương lai,thoi-trang-tuong-lai,2025-07-22 10:25:20,2025-07-22 10:25:20
7,Tập đoàn Tài chính Zenith,tap-doan-tai-chinh-zenith,2025-07-22 10:25:20,2025-07-22 10:25:20
8,Thực phẩm Sức sống,thuc-pham-suc-song,2025-07-22 10:25:20,2025-07-22 10:25:20
9,Phòng thí nghiệm Kỹ thuật số,phong-thi-nghiem-ky-thuat-so,2025-07-22 10:25:20,2025-07-22 10:25:20
10,Vườn Ốc đảo Đô thị,vuon-oc-dao-do-thi,2025-07-22 10:25:20,2025-07-22 10:25:20
11,Giải pháp Năng lượng Xanh,giai-phap-nang-luong-xanh,2025-07-22 10:25:20,2025-07-22 10:25:20
12,Xưởng Sáng tạo Vô cực,xuong-sang-tao-vo-cuc,2025-07-22 10:25:20,2025-07-22 10:25:20
13,Tập đoàn Hậu cần Ưu việt,tap-doan-hau-can-uu-viet,2025-07-22 10:25:20,2025-07-22 10:25:20
14,Tự động hóa Nhà thông minh,tu-dong-hoa-nha-thong-minh,2025-07-22 10:25:20,2025-07-22 10:25:20
15,Công ty Ẩm thực Tuyệt hảo,cong-ty-am-thuc-tuyet-hao,2025-07-22 10:25:20,2025-07-22 10:25:20
16,Công ty Du lịch Thám hiểm,cong-ty-du-lich-tham-hiem,2025-07-22 10:25:20,2025-07-22 10:25:20
17,Hệ thống Lọc nước Tinh khiết,he-thong-loc-nuoc-tinh-khiet,2025-07-22 10:25:20,2025-07-22 10:25:20
18,Học viện Học tập Chú tâm,hoc-vien-hoc-tap-chu-tam,2025-07-22 10:25:20,2025-07-22 10:25:20
19,Sản phẩm Làm đẹp Rạng rỡ,san-pham-lam-dep-rang-ro,2025-07-22 10:25:20,2025-07-22 10:25:20
20,Kỹ thuật Sáng tạo,ky-thuat-sang-tao,2025-07-22 10:25:20,2025-07-22 10:25:20
21,Dịch vụ Giao hàng Nhanh,dich-vu-giao-hang-nhanh,2025-07-22 10:25:20,2025-07-22 10:25:20
22,Nội thất Không gian Thanh bình,noi-that-khong-gian-thanh-binh,2025-07-22 10:25:20,2025-07-22 10:25:20
23,Phòng thí nghiệm Nghiên cứu Lượng tử,phong-thi-nghiem-nghien-cuu-luong-tu,2025-07-22 10:25:20,2025-07-22 10:25:20
24,Đồ dùng Gia đình Thu hoạch đồ dùng gia đình thu hoạch,2025-07-22 10:25:20,2025-07-22 10:25:20
```

Bên cạnh đó, Genmini được sử dụng để viết các Rules và Randomized range cho từng thuộc tính

id	trang dan , từ 1 - 100
First Name	Họ phổ biến
Last Name	Tên phổ biến
Address	Địa chỉ gồm tên và số
City	Các thành phố nổi tiếng trên thế giới
State	Áp dụng nếu quốc gia đó có hệ thống bang, ngược lại thì NULL
Country	Khớp với City
Postal code	Ứng với mã code của từng quốc gia
Phone	Là chuỗi số gồm 11-12 ký tự
DOB	Ngày sinh trong quá khứ (trước ngày hiện tại)
Email	Theo định dạng abc@example.com
Password	Chuỗi được mã hóa
Role	user



DỰA THEO BẢNG MÔ TẢ MIỄN GIÁ TRỊ USERS NÀY BẠN
HÃY VIẾT TIẾP VỚI BẢNG PRODUCTS


idnamedescriptionstockpriceis_location_offeris_rentalbr
and_idcategory_idproduct_image_idcreated_atupdated_
at1Screwdriver SetA versatile set of screwdrivers for
various tasks.15025.99101112025-07-19 0:00:002025-
07-20 0:00:002Adjustable WrenchHeavy-duty
adjustable wrench for plumbing and general
repairs.8018.5102122025-07-19 0:00:002025-07-20
0:00:003Cordless DrillPowerful cordless drill with
multiple speed settings.5079.99103232025-07-19
0:00:002025-07-20 0:00:00

Ta thực hiện tương tự với lần lượt 6 bảng,

Gemini gen ra kết quả và link Google sheets.

Invoice

Field	Rules/Ranges Ngắn gọn
id	Tăng dần, 1 - 5000 (hoặc số lượng hóa đơn dự kiến)
user_id	ID User hợp lệ (tham chiếu bảng Users)
invoice_date	Ngày giờ trong quá khứ hoặc hiện tại
invoice_number	Chuỗi duy nhất, số hóa đơn
billing_address	Địa chỉ thanh toán (tương tự địa chỉ User)
billing_city	Thành phố thanh toán (tương tự City của User)
billing_state	Bang/Tỉnh thanh toán (tương tự State của User)
billing_country	Quốc gia thanh toán (tương tự Country của User)
billing_postcode	Mã bưu chính thanh toán (tương tự Postal code của User)
total	Số thực, bằng unit_price của sản phẩm trong Invoice Item tương ứng
payment_method	Phương thức thanh toán (ví dụ: "Cash on Delivery", "Bank transfer", "Credit Card")
payment_account_name	Tên tài khoản thanh toán (nếu có)
payment_account_number	Số tài khoản thanh toán (nếu có)
status	Trạng thái hóa đơn (ví dụ: "AWAITING_FULFILLMENT", "COMPLETED", "CANCELLED")
status_message	Tin nhắn trạng thái bổ sung hoặc NULL
created_at	Ngày giờ quá khứ
updated_at	Ngày giờ sau Created_at hoặc NULL

 Export to Sheets

2. Python (thư viện Fakers)

```
def generate_contact_replies(num_replies):
    """Generates contact reply data."""
    if not user_ids:
        print("Please generate users first to create contact replies.")
        return [], []

    headers = [
        'id', 'contact_id', 'user_id', 'reply_message', 'replied_at', 'created_at', 'updated_at'
    ]
    data = []

    # Danh sách 5 phản hồi mặc định
    default_replies = [
        "Cảm ơn bạn đã liên hệ! Chúng tôi đã nhận được yêu cầu của bạn và sẽ phản hồi sớm nhất.",
        "Xin lỗi vì sự chậm trễ. Vấn đề của bạn đang được xử lý bởi đội ngũ hỗ trợ kỹ thuật của chúng tôi.",
        "Chúng tôi đã cung cấp giải pháp cho vấn đề của bạn. Vui lòng kiểm tra email của bạn để biết chi tiết.",
        "Để hiểu rõ hơn về vấn đề, bạn có thể cung cấp thêm thông tin không?",
        "Chúng tôi đang điều tra vấn đề này và sẽ cập nhật cho bạn ngay khi có thông tin mới."
    ]

    dummy_contact_ids = list(range(1, 101)) # Giả định có 100 contact_id

    for i in range(1, num_replies + 1):
        reply_id = i
        contact_id = random.choice(dummy_contact_ids)
        user_id = random.choice(user_ids)

        # Chọn ngẫu nhiên một phản hồi từ danh sách mặc định
        reply_message = random.choice(default_replies)

        replied_at = datetime.now().strftime('%Y-%m-%d %H:%M:%S')
        created_at = replied_at
        updated_at = 'NULL'

        data.append([
            reply_id, contact_id, user_id, reply_message, replied_at, created_at, updated_at
        ])
    return data, headers
```

Ý tưởng: Thay vì tạo ra các tin nhắn phản hồi mới hoàn toàn ngẫu nhiên mỗi lần, chúng ta định nghĩa trước một tập hợp các phản hồi phổ biến (ví dụ: 5 câu). Khi cần một tin nhắn phản hồi, hệ thống sẽ **chọn ngẫu nhiên một trong các câu đã được định nghĩa sẵn** đó. Điều này giúp đảm bảo các phản hồi có ý nghĩa và nhất quán hơn, đồng thời vẫn giữ được sự đa dạng ở mức độ nhất định.

Quy trình:

```
PS D:\Nam_3\HK3\testing\homework\HW3\script> python homework.py
Enter table name (users, category, brand, product, product_image, invoice, invoice_item, contact_reply): contact_reply
Enter number of records to generate: 2
Generating 10 users first to ensure contact reply data can be created...
id,contact_id,user_id,reply_message,replied_at,created_at,updated_at
1,53,9,"Để hiểu rõ hơn về vấn đề, bạn có thể cung cấp thêm thông tin không?",2025-07-22 12:48:26,2025-07-22 12:48:26,NULL
2,38,6,Cảm ơn bạn đã liên hệ! Chúng tôi đã nhận được yêu cầu của bạn và sẽ phản hồi sớm nhất.,2025-07-22 12:48:26,2025-07-22 12:48:26,NULL

Generated 2 records for the 'contact_reply' table.
PS D:\Nam_3\HK3\testing\homework\HW3\script>
```

Chọn bảng muốn sinh dữ liệu
Sau đó chọn số dòng muốn sinh
Kết quả được hiển thị

Assessment Criteria

Criteria	Description	Max Points
2 Tables Selection	2 important tables selected	1.0/1.0
Sample data	All data must be meaningful	2.0/2.0
Data generation report	Report is clear, traceable, professional, with self-assessment	1.0/1.0
2 Scenario Selection	2 important scenario selected	1.0/1.0
Scenario testing	Correct and complete scenario identification	2.0/2.0
Use of AI Tools	Prompt transparency, critical validation, added value	1.0/1.0
Test Execution	All designed test cases executed, results logged	0.5/0.5

Bug Reporting	Clear and complete bug report(s), if applicable	0.5/0.5
Scenario testing report	Report is clear, traceable, professional, with self-assessment	0.5/1.0
Total		9.5/1.0 points