TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÀI TẬP LỚN MÔN HỆ CƠ SỞ DỮ LIỆU**

*Người hướng dẫn*: TS DOÃN XUÂN THANH

*Người thực hiện*: **BÙI PHƯƠNG NAM – 522H0002**

Lớp **: 22H50201**

Khoá  **: K26**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

LỜI CẢM ƠN

Lời đầu tiên, em xin gửi lời cảm ơn đến Thầy Doãn Xuân Thanh. Trong quá trình học cũng như là làm bài tiểu luận, nhờ có sự giúp đỡ tận tình của Thầy, Thầy đã truyền đạt cho chúng em những kiến thức bổ ích. Thầy đã hướng dẫn chúng em nhiệt tình, nhờ đó mà chúng em có thể hoàn thành bài tiểu luận này.HỆ CƠ SỞ DỮ LIỆU là môn học thú vị và nhưng cũng rất phức tạp. Những kiến thức và kỹ năng về môn học này của em vẫn còn nhiều hạn chế. Do đó, bài tập của em khó tránh khỏi sai sót. Rất mong nhận được những góp ý của Thầy về bài tập này.

Một lần nữa, em xin chân thành cảm ơn!

**BÀI TẬP ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là sản phẩm đồ án của riêng tôi và được sự hướng dẫn của TS DOÃN XUÂN THANH. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày tháng năm*

*Tác giả*

*Bùi Phương Nam*

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

**Phần xác nhận của GV hướng dẫn**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

**Phần đánh giá của GV chấm bài**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

*TÓM TẮT*

Trong bài tập này, chúng ta sẽ tiếp cận và giải quyết một số yêu cầu liên quan đến Hệ Cơ sở dữ liệu. Bài tập được thiết kế để kiểm tra và phát triển kỹ năng về thiết kế cơ sở dữ liệu, sử dụng SQL, và lập trình bằng Java hoặc Python. Bài tập có 2 phần

I. Phần 1: Thiết kế cơ sở dữ liệu (6 Điểm)

Sinh viên sẽ tự viết đặc tả cho một nghiệp vụ có các mối quan hệ khác nhau. Các mối quan hệ sẽ được đánh giá dựa trên loại hình và cấu trúc của chúng. Sau đó, sinh viên sẽ vẽ mô hình ERD và chuyển đổi nó sang mô hình quan hệ.

II. Phần 2: Triển khai một số chức năng bằng Java hoặc Python

Sinh viên sẽ sử dụng ngôn ngữ lập trình Java hoặc Python để triển khai chức năng xử lý đầu vào và đầu ra theo yêu cầu. Điều này bao gồm xử lý mô hình ERD từ một tệp đầu vào và tạo các bảng dữ liệu, mối quan hệ. Ngoài ra, sinh viên cũng sẽ triển khai lược đồ CSDL và xử lý các yêu cầu liên quan đến thuộc tính và phụ thuộc hàm.

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc387692905)

[PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN i](#_Toc387692906)

[TÓM TẮT i](#_Toc387692907)

[MỤC LỤC 1](#_Toc387692908)

[DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ 1](#_Toc387692909)

[PHẦN 1 1](#_Toc387692910)

1. [Mô tả nghiệp vụ 1](#_Toc387692911)

2. Mô hình ERD được xây dựng từ mô tả

3. Xây dựng mô hình quan hệ từ ERD.

4. Sử dụng câu lệnh SQL để thực hiện các thao tác

4.1 Tạo các bảng cho mô hình quan hệ

4.2 Tạo ít nhất 2 function để thêm dữ liệu vào các bảng

4.3 Tạo trigger kiểm tra ràng buộc các bảng

PHẦN 2

[2.1 Chức năng 1 1](#_Toc387692918)

[2.2 Chức năng 2 1](#_Toc387692919)

Danh sách bài đã làm và bài chưa làm:

1. Bài đã làm

Phần 1:

* Xây dựng một mô tả nghiệp vụ
* Các mối quan hệ
* Vẽ mô hình ERD cho đặc tả (1 điểm)
* Chuyển mô hình ERD sang mô hình quan hệ
* Sử dụng câu lệnh SQL để thực hiện các thao tác sau:
  1. Tạo các bảng cho mô hình quan hệ trên
  2. Tạo ít nhất 2 function để thêm dữ liệu vào các bảng trên

Phần 2:

Input: File Input1.txt chứa các thực thể và mối quan hệ giữa các thực thể của mô hình ERD. Output: File Output1.txt chứa các bảng dữ liệu và mối quan hệ giữa các bảng. Định dạng của file Input1.txt và Output1.txt do sinh viên tự thiết lập. (2 điểm)

1. Bài chưa làm

Phần 1:

* Tạo ít nhất một trigger để kiểm tra ràng buộc khoá ngoại, ràng buộc miền giá trị

Phần 2:

Input: File Input2.txt chứa lược đồ CSDL bao gồm các bảng, các thuộc tính và các phụ thuộc hàm. Output: File Output2.txt chứa bao đóng của tập thuộ\c tính X (Tập thuộc tính X do người dùng truyền vào); các khoá của lược đồ quan hệ. Định dạng của file Input2.txt và Output2.txt do sinh viên tự thiết lập.

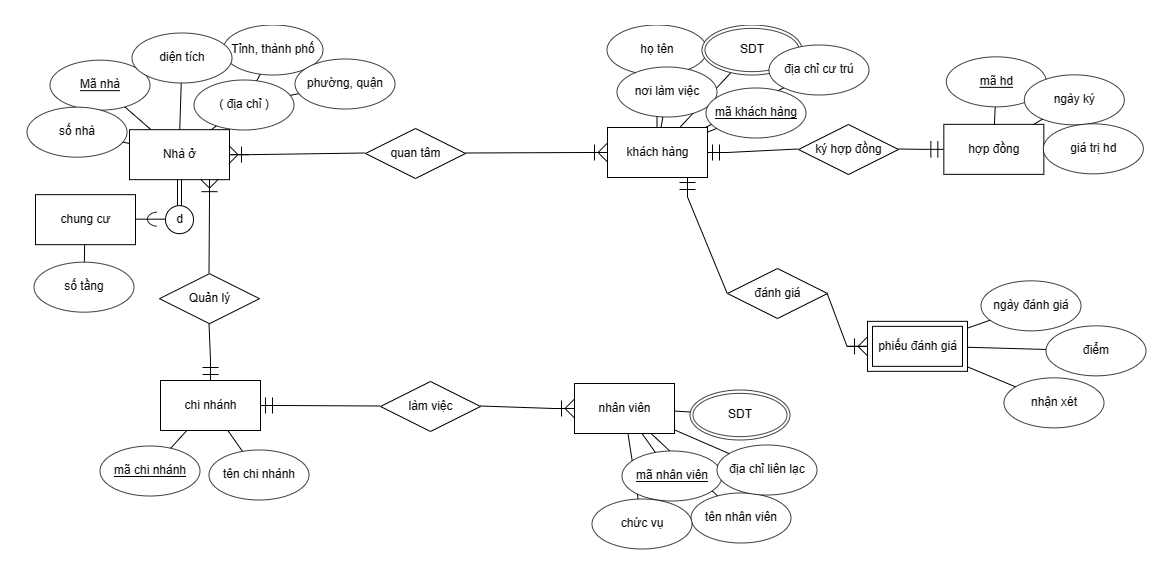
1. Input: File Input2.txt chứa lược đồ CSDL bao gồm các bảng, các thuộc tính và các phụ thuộc hàm. Output: File Output2.txt chứa bao đóng của tập thuộ\c tính X (Tập thuộc tính X do người dùng truyền vào); các khoá của lược đồ quan hệ. Định dạng của file Input2.txt và Output2.txt do sinh viên tự thiết lập. (2 điểm)

Phần I

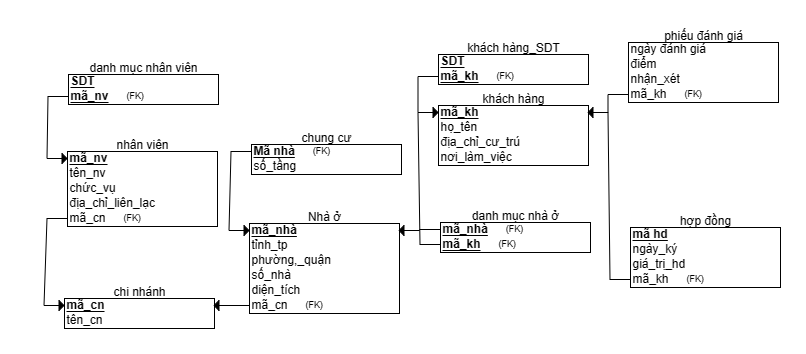
1. Xây dựng một mô tả nghiệp vụ

Một công ty bất động sản chuyên kinh doanh về nhà ở. Nhà ở có nhiều loại nhưng công ty chỉ cung cấp 2 loại chính là chung cư và nhà ở riêng lẻ. Thông tin về nhà ở bao gồm mã nhà , địa chỉ (ghi rõ thuộc phường/quận nào, tỉnh/tp nào) , số nhà , diện tích. Đối với loại hình chung cư thì có thêm thông tin về số tầng. Các chi nhánh của công ty sẽ chịu trách nhiệm quản lý việc giao dịch các nhà ở trên. Một nhà ở được quản lý bởi một chi nhánh nhưng một chi nhánh có thể quản lý nhiều nhà ở. Thông tin về chi nhánh gồm có mã chi nhánh và tên chi nhánh. Một chi nhánh sẽ có nhiều nhân viên nhưng một nhân viên chỉ thuộc một chi nhánh. Một nhân viên sẽ có thông tin: tên, ngày sinh, sdt, chức vụ, địa chỉ liên lạc. Mỗi khách hàng có thể quan tâm nhiều sản phẩm của công ty để lựa chọn và nhiều khách hàng có thể quan tâm cùng một sản phẩm. Mỗi khách hàng khi giao dịch sẽ phải làm một hợp đồng. Thông tin về hợp đồng bao gồm mã hợp đồng, ngày ký, giá trị hợp đồng. Khách hàng làm hợp đồng sẽ được ghi nhận một số thông tin bao gồm họ tên, SDT, địa chỉ cư chú, nơi làm việc, tài chính, lịch sử hoạt động, mã khách hàng. Ngoài ra công ty sẽ ghi nhận đánh giá của khách hàng về sản phẩm thông qua phiếu đánh giá (không bắt buộc). Thông tin về phiếu đánh giá gồm điểm số (từ 1 đến 5), nhận xét, ngày đánh giá. Khách hàng có thể đánh giá nhiều lần

1. Xây dựng mô hình ERD dựa trên mô tả đã cho



1. Chuyển mô hình erd thành mô hình quan hệ



1. Sử dụng câu lệnh SQL để thực hiện các thao tác dựa trên mô hình quan hệ
   1. Tạo bảng cho mô hình quan hệ

Dựa vào mô hình quan hệ dữ liệu trên, ta sẽ có các bảng sau:

* nhà\_ở(mã\_nhà, tỉnh\_tp, phường\_quận, số\_nhà, diện\_tích , mã\_cn)
* Danh\_mục\_nhà\_ở(**mã\_nhà**, mã\_kh)
* chi\_nhánh(**mã\_cn**, tên\_cn)
* nhân\_viên(**mã\_nv**,tên\_nv, chức\_vụ, địa\_chỉ\_liên\_lạc, mã\_cn)
* danh\_mục\_nhân\_viên(**mã\_nv**, SDT)
* khách\_hàng(mã\_kh, họ\_tên, nơi\_làm\_việc, địa\_chỉ\_cư\_trú)
* danh\_mục\_khách\_hàng(mã\_kh, SDT)
* chung\_cư(**mã\_nhà**, số\_tầng)
* hợp\_đồng(**mã\_hd**, ngày\_ký, giá\_trị\_hd, mã\_kh)
* phiếu\_dánh\_giá(ngày\_dg, điểm, nhận\_xét, mã\_kh)

4.2 Tạo ít nhất 2 function để thêm dữ liệu vào các bảng

Thực tế, trong sql server, một function sẽ không được phép thêm dữ liệu vào bảng vì đây là công việc của procedure và trigger. Một function sẽ dùng để phát sinh khóa chính cho đối tượng mới khi ta thêm đối tượng đó vào bảng . Một số khóa chính của các bảng sẽ được phát sinh như sau:

* Đối với bảng khách\_hàng và danh\_mục\_khách\_hàng thì khóa chính mã\_kh sẽ được phát sinh theo nguyên tắc: KH + STT.

Trong đó KH là viết tắt của “khách hàng” và STT là số thứ tự của khách hàng

Mã khách hàng có tối đa 10 ký tự. VD: KH152 (nghĩa là khách hàng đã ký hợp đồng thứ 152)

* Đối với bảng hợp\_đồng thì mã\_hd sẽ được phát sinh theo quy tắc:

HD + STT + mã quý. Trong đó:

* + HD là viết tắt của hợp đồng
  + STT là số thứ tự của khách hàng ký hợp đồng, lấy từ bảng khách hàng
  + mã quý sẽ nhận một trong các giá trị (Q1, Q2, Q3, Q4) tùy thuộc khách hàng ký hợp đồng trong quý nào trong năm

VD: HD150Q3 (nghĩa là hợp đồng của khách hàng thứ 150, được ký vào quý 3). Mã hd sẽ có tối đa 10 ký tự

* Đối với bảng chi\_nhánh thì mã\_cn sẽ được phát sinh theo quy tắc:

DVC\_mã quận\_mã thành phố\_STT. Trong đó:

* + DVC là mã công ty
  + STT là số thứ tự chi nhánh
  + Mã thành phố và quận được tìm trên web cổng thông tin của chính phủ

VD: DVC\_778\_79\_0124 (nghĩa là chi nhánh thứ 0124 của công ty DVC ngụ tại quận 7 (778), thuộc tp.HCM (79) )

* Đối với bảng nhân\_viên và danh\_mục\_nhân\_viên thì mã\_nv sẽ được phát sinh theo quy tắc NV + STT . Trong đó:
  + NV là viết tắt của “nhân viên”
  + STT là số thứ tự

VD: NV0145 . Mã nhân viên có tối đa 8 ký tự

* Đối với bảng nhà ở và danh mục nhà ở thì mã\_nhà sẽ được phát sinh theo quy tắc: mã tp /mã quận/mã phường /STT nhà ở

VD: 79/27481/778/0134 , nghĩa là nhà ở thứ 0134 nằm tại quận 7 (778), phường Tân Quy (27481), thuộc tp.HCM (74) . Mã nhà có tối đa 10 ký tự

Ứng với mỗi quy tắc phát sinh, ta sẽ tạo một function để sinh ra các khóa chính, các khóa còn lại của bảng sẽ được thêm vào nhờ các procedure

4.3 Tạo ít nhất một trigger để kiểm tra ràng buộc khoá ngoại, ràng buộc miền giá trị.

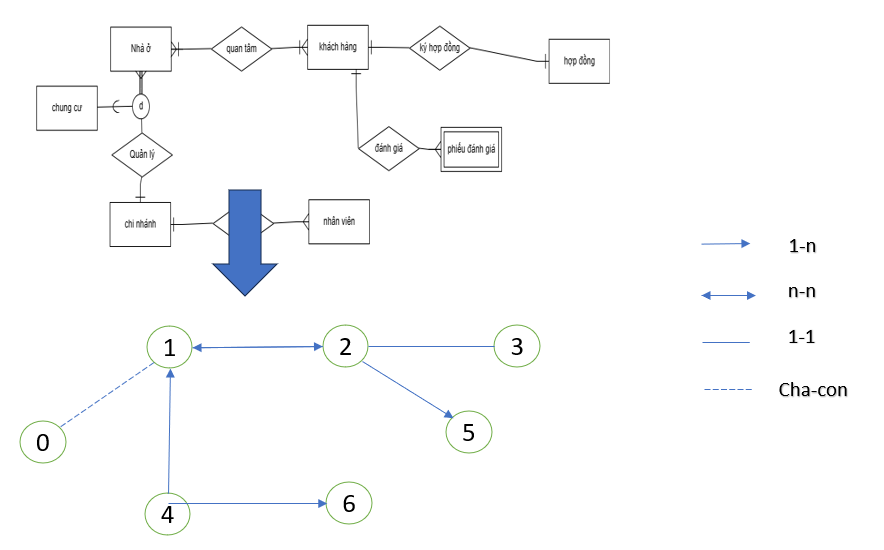
**PHẦN 2:**

Sinh viên sử dụng ngôn ngữ Java hoặc Python để cài đặt chức năng sau:

1. Input: File Input1.txt chứa các thực thể và mối quan hệ giữa các thực thể của mô hình ERD. Output: File Output1.txt chứa các bảng dữ liệu và mối quan hệ giữa các bảng. Định dạng của file Input1.txt và Output1.txt do sinh viên tự thiết lập.

**Ý tưởng**

Ta sẽ lấy mô hình quan hệ ở Phần 1 và coi như một đồ thị mà trong đó, mỗi đỉnh là một thực thể. Các cạnh sẽ là các quan hệ giữa các thực thể, song đồ thị sẽ được tùy biến sao cho phù hợp với mô hình



Trong đồ thị trên, các đỉnh là các thực thể được định danh bằng số tự nhiên 0, 1, 2, 3,…. Những đoạn nối 2 đỉnh có một mũi tên biểu thị quan hệ 1-n, những đoạn có 2 mũi tên ở 2 đầu sẽ biểu thị quan hệ n-n, những đoạn không có mũi tên sẽ biểu thị quan hệ 1-1, còn những đoạn nét đứt sẽ biểu thị quan hệ cha và con.

Lưu ý: ta phân biệt quan hệ 1-n với n-1 để xác định xem thực thể nào nằm ở phía n và thực thể nào nằm ở phía 1. Đối với cặp thực thể cha – con, thực thể có định danh lớn hơn sẽ là thực thể cha. Nếu một cha có nhiều con, định danh của cha sẽ là lớn nhất, định danh của các con sẽ là ngẫu nhiên nhưng nhỏ hơn định danh của cha

Trong file input1.txt, ta có thể biểu diễn đồ thị trên dưới dạng danh sách cạnh với mỗi cạnh là một thực thể có 3 thuộc tính (u,v,w) đại diện cho đỉnh u, đỉnh v, và w là quan hệ giữa u và v . Từ đồ thị, ta có thể xây dựng các bảng và đưa chúng vào Output1.txt

Vì thực thể có định danh nên ta sẽ dùng hashMap để lập danh sách các cặp key – value tương ứng với cặp Tên thực thể - định danh. VD: khách hàng – 0, chi nhánh – 1, hợp đồng – 2 ,… Ngoài ra, với hashMap, ta còn có thể tìm các cặp thực thể có quan hệ với nhau

Thực thi với ngôn ngữ Java (Nhưng không dùng được tiếng Việt)

**Thực thi xử lý**

* Khởi tạo hashMap lưu trữ các cặp key – value. Trong đó key sẽ có kiểu int, đại diện cho định danh, value có kiểu String, là thực thể và các thuộc tính của nó. Trong mỗi thực thể, nếu thuộc tính là khóa chính thì được đánh dấu \*. Đối với thuộc tính phức thì thay nó bằng các thuộc tính con.
* Duyệt qua danh sách cạnh, với mỗi cạnh (u,v,w), tiến hành:
  + Ánh xạ cặp đỉnh u, v vào hashMap 🡪 xác định cặp thực thể
  + Kiểm tra w, nếu:
    - w là ‘1-n’ hoặc ‘n-1’ thì đưa khóa chính của thực thể phía 1 vào làm khoá ngoại của của thực thể phía n. (2 bước kiểm tra)
    - w là ‘n-n’ thì tạo một bảng mới chứa khóa chính của cặp thực thể đó. Các khóa chính đó cũng đóng vai trò làm khóa ngoại. Tên bảng mới tạm thời được ghép từ 2 tên của cặp thực thể
    - w là '1-1’ thì khoá chính của 1 trong 2 thực thể sẽ được copy vào làm khoá ngoại của quan hệ kia. Tuy nhiên, nên ưu tiên đưa khóa chính của thực thể có định danh thấp hơn sang làm khóa ngoại của thực thể có định danh cao hơn (ví dụ 2 < 3 thì lấy khóa chính của 2 làm khóa ngoại của 3)
    - w là ‘cha-con’ , thì so sánh định danh của chúng để xác định đâu là cha và đâu là con, sau đó copy toàn bộ khóa của thực thể cha đưa cho thực thể con
* Sau khi cập nhật các thực thể, thì lưu chúng vào hashMap cho tới khi hết danh sách liên kết
* Ghi dữ liệu trong HashMap vào file Output.txt

Lưu ý: quá trình cập nhật thực thể sẽ khó khăn do gặp vấn đề về xử lý chuỗi. Ta sẽ áp dụng các thao tác với biểu thức chính quy (regex) của java để xử lý.

Input:

0 1 cha-con

1 2 n-n

1 4 n-1

4 6 1-n

2 3 1-1

2 5 1-n

Output:

Apartment(\*House\_id, province\_city, ward\_district, number, area, floor\_number)

House(\*House\_id, province\_city, ward\_district, number, area)

Customer(\*customer\_id, full\_name, work\_place, address, phone\_number)

Contract(\*contract\_id, sign\_date, contract\_available, \*customer\_id)

Branch(\*branch\_id, branch\_name, \*House\_id)

Vote(vote\_date, score, assesst, \*customer\_id)

Employee(employee\_id, full\_name, position, address, phone\_number, \*branch\_id)

House\_Customer(\*House\_id, \*customer\_id)

1. Input: File Input2.txt chứa lược đồ CSDL bao gồm các bảng, các thuộc tính và các phụ thuộc hàm. Output: File Output2.txt chứa bao đóng của tập thuộ\c tính X (Tập thuộc tính X do người dùng truyền vào); các khoá của lược đồ quan hệ. Định dạng của file Input2.txt và Output2.txt do sinh viên tự thiết lập.