

**TRƯỜNG ĐẠI HỌC MỞ THÀNH PHỐ HỒ CHÍ MINH**  
**KHOA CÔNG NGHỆ THÔNG TIN**

## **BÁO CÁO TỔNG KẾT**

**ĐỀ TÀI NGHIÊN CỨU KHOA HỌC SINH VIÊN**

**XÂY DỰNG DATA WAREHOUSE LƯU TRỮ DỮ LIỆU  
CHO HOẠT ĐỘNG PHÂN PHỐI NƯỚC GIẢI KHÁT**

**742**

**Chủ nhiệm đề tài : Nguyễn Đặng Phú Mẫn**

**Khoa : Công nghệ thông tin**

**Các thành viên : Phạm Công Thuận**

**Người hướng dẫn : Võ Thị Kim Anh**

## MỤC LỤC

<b>MỤC LỤC.....</b>	<b>1</b>
<b>LỜI MỞ ĐẦU.....</b>	<b>3</b>
<b>CHƯƠNG 1. NGHIÊN CỨU CƠ SỞ LÝ THUYẾT.....</b>	<b>4</b>
<b>1. KHÁI NIỆM DATA WAREHOUSE.....</b>	<b>4</b>
1.1. ĐỊNH NGHĨA KHO DỮ LIỆU - DATA WAREHOUSE.....	4
1.2. ĐẶC TÍNH KHO DỮ LIỆU - DATA WAREHOUSE.....	4
1.3. KIẾN TRÚC KHO DỮ LIỆU - DATA WAREHOUSE.....	5
<b>2. TÌM HIỂU HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU SQL SERVER.....</b>	<b>7</b>
2.1. KHÁI QUÁT VỀ SQL SERVER.....	7
2.2. ỨNG DỤNG SQL SERVER CHO DATA WAREHOUSE.....	7
<b>3. TÌM HIỂU CÔNG CỤ SQL SERVER INTEGRATION SERVICES.....</b>	<b>8</b>
3.1. KHÁI QUÁT VỀ SQL SERVER INTEGRATION SERVICES (SSIS).....	8
3.2. CHỨC NĂNG CỦA SSIS.....	8
3.3. THÀNH PHẦN CHÍNH CỦA SSIS.....	9
<b>4. TÌM HIỂU PHẦN MỀM POWER BI.....</b>	<b>10</b>
4.1. KHÁI QUÁT VỀ POWER BI.....	10
4.2. ỨNG DỤNG POWER BI CHO VIỆC TRỰC QUAN HÓA DỮ LIỆU....	10
<b>CHƯƠNG 2. THỰC NGHIỆM XÂY DỰNG DATA WAREHOUSE.....</b>	<b>11</b>
<b>1. QUY TRÌNH TIẾN HÀNH THỰC NGHIỆM.....</b>	<b>11</b>
<b>2. THIẾT KẾ CƠ SỞ DỮ LIỆU GỐC.....</b>	<b>11</b>
<b>3. THIẾT KẾ CƠ SỞ DỮ LIỆU STAGING.....</b>	<b>15</b>
<b>4. THIẾT KẾ DATA WAREHOUSE THEO MÔ HÌNH STAR SCHEMA.....</b>	<b>15</b>
<b>5. THỰC HIỆN EXTRACT - TRANSFORM - LOAD (ETL) BẰNG SQL SERVER INTEGRATION SERVICES (SSIS).....</b>	<b>16</b>
5.1. CẤU HÌNH CONNECTION MANAGER.....	16
5.2. THỰC HIỆN SAO CHÉP DỮ LIỆU TỪ CƠ SỞ DỮ LIỆU BEVERAGE SANG BEVERAGESTAGE.....	19
5.3. CHUYỂN ĐỔI VÀ TẢI DỮ LIỆU TỪ BEVERAGESTAGE VÀO BEVERAGEDW.....	26
5.3.1. TẢI DỮ LIỆU TỪ KHÁCH HÀNG VÀO DATA WAREHOUSE... 26	
5.3.2. TẢI DỮ LIỆU SẢN PHẨM VÀO DATA WAREHOUSE.....	30
5.3.3. TẢI DỮ LIỆU BÁN HÀNG VÀO DATAWARE HOUSE.....	36
5.4. TRỰC QUAN HÓA DỮ LIỆU BẰNG POWER BI.....	41
5.4.1. LẤY DỮ LIỆU TỪ DATA WAREHOUSE.....	41
5.4.2. TẠO DASHBOARD.....	43
5.4.3. TẠO SALES REPORT.....	47
<b>KẾT LUẬN.....</b>	<b>53</b>
<b>❖ TÀI LIỆU THAM KHẢO.....</b>	<b>54</b>

## LỜI MỞ ĐẦU

Trong bối cảnh ngày nay, các loại sản phẩm về nước giải khát ngày càng gia tăng và trở nên đa dạng hơn theo nhu cầu của người tiêu dùng kèm theo sự cạnh tranh gay gắt đòi hỏi các doanh nghiệp phải ứng dụng các công nghệ hiện đại để quản lý dữ liệu doanh nghiệp. Do đó, việc tối ưu hóa quá trình lưu trữ và xử lý dữ liệu cho hoạt động phân phối nước giải khát trở thành một thách thức với các doanh nghiệp hoạt động trong lĩnh vực thức uống.

Đề tài nghiên cứu này xác định cách thức xây dựng và tối ưu hóa một Data Warehouse lưu trữ dữ liệu phân phối nước giải khát. Mục tiêu của đề tài chính là tạo ra một hệ thống hoàn chỉnh để thu thập, lưu trữ và phân tích các dữ liệu liên quan đến việc phân phối các sản phẩm này.

Đối tượng nghiên cứu của đề tài bao gồm những doanh nghiệp hoạt động trong ngành sản xuất và phân phối nước giải khát, cụ thể hơn là các nhân sự quản lý dữ liệu có liên quan đến việc xây dựng và quản lý hệ thống Data Warehouse. Đề tài tập trung nghiên cứu các giai đoạn cần thiết để xây dựng hệ thống Data Warehouse, từ việc thu thập dữ liệu, thiết kế cơ sở dữ liệu, triển khai hệ thống cho đến việc đánh giá hiệu suất và phương hướng phát triển. Phương pháp nghiên cứu sẽ sử dụng một kết hợp các phương pháp hành động và phân tích để đảm bảo sự hiệu quả và tính khả thi của giải pháp được đề xuất.

## **CHƯƠNG 1. NGHIÊN CỨU CƠ SỞ LÝ THUYẾT**

### **1. KHÁI NIỆM DATA WAREHOUSE**

#### **1.1. ĐỊNH NGHĨA KHO DỮ LIỆU - DATA WAREHOUSE**

Kho dữ liệu trước đây được định nghĩa là một tập hợp dữ liệu được dùng để hỗ trợ quy trình ra quyết định. Qua quá trình phát triển sau này, Kho dữ liệu được hiểu là một môi trường thông tin (information environment) chứa dữ liệu lớn của doanh nghiệp.

Các chức năng của Kho dữ liệu bao gồm:

- Cung cấp góc nhìn toàn diện về hoạt động doanh nghiệp: Trong trường hợp doanh nghiệp có nhiều mảng kinh doanh được quản lý bởi các hệ thống khác nhau thì Kho dữ liệu là nơi đã tập hợp (tích hợp) được mọi thông tin về các mảng nghiệp vụ đó để cung cấp một góc nhìn toàn diện. Việc này còn cung cấp khả năng đánh giá các mảng nghiệp vụ khác nhau để phân tích sự tương quan giữa chúng [1].
- Cung cấp đầy đủ thông tin hiện tại và trong quá khứ của doanh nghiệp: Dữ liệu được lưu không chỉ là dữ liệu thô lấy từ các hệ thống tác nghiệp mà nó đã được tổng hợp, tính toán thành các độ đo có ý nghĩa phân tích phục vụ cho việc khai thác, sử dụng cho việc hỗ trợ ra quyết định chiến lược [1].
- Cung cấp dữ liệu chi tiết theo nhu cầu mà không cần truy xuất các hệ thống tác nghiệp: Trong một số trường hợp, nhu cầu phân tích dữ liệu chỉ ở mức giao dịch thì nó đã được được lưu sẵn ở Kho dữ liệu.
- Đảm bảo thông tin trong Kho dữ liệu có tính nhất quán: Một trường dữ liệu có thể được biểu diễn bằng nhiều kiểu khác nhau. Giả sử Tên khách hàng có hệ thống tách họ và tên thành 2 trường khác nhau, có hệ thống chỉ lưu trong 1 trường. Sau khi tổng hợp về Kho dữ liệu, trường này sẽ được chuyển về một dạng thống nhất [1].
- Là nguồn thông tin chiến lược mềm dẻo và có tính tương tác: Người dùng có thể truy xuất nhiều thông tin khác nhau của một đối tượng và thực hiện được nhiều thao tác với đối tượng dữ liệu thay vì trả lại một danh sách tĩnh. Tính tương tác còn thể hiện ở việc người dùng có thể áp dụng các thao tác phân tích dữ liệu, chẳng hạn như nền tảng phân tích dữ liệu SQL Server Analysis Services giúp người dùng có thể tương tác để phân tích dữ liệu.

## 1.2. ĐẶC TÍNH KHO DỮ LIỆU - DATA WAREHOUSE

Kho dữ liệu - Data Warehouse bao gồm 4 đặc tính:

- Đặc tính Subject-oriented (Hướng chủ đề): Kho dữ liệu được thiết kế theo chủ đề, tập trung vào các vấn đề liên quan tới quá trình kinh doanh của doanh nghiệp. Điều này giúp cho việc phân tích dữ liệu dễ dàng hơn khi người dùng có thể nắm được mọi thông tin về một chủ đề cụ thể một cách nhanh chóng [2].
- Đặc tính Integrated (Nguồn đa dạng): Kho dữ liệu tích hợp dữ liệu từ nhiều nguồn khác nhau (ứng dụng, website,...), đảm bảo tính nhất quán và đúng đắn của dữ liệu.
- Đặc tính Time variant (Dữ liệu được gán thời gian): Kho dữ liệu chứa dữ liệu quá khứ và dữ liệu cập nhật, giúp cho người dùng có thể phân tích dữ liệu theo thời gian.
- Đặc tính Non-volatile (Dữ liệu không biến đổi): Dữ liệu trong kho không bị thay đổi hay xóa bỏ, ngoại trừ các trường hợp đặc biệt như bảo trì hoặc cập nhật dữ liệu. Điều này giúp cho kho dữ liệu trở nên đáng tin cậy và bảo mật hơn [2].



### 4 ĐẶC TÍNH QUAN TRỌNG CỦA KHO DỮ LIỆU



Hình 1.1.2.1. Các đặc tính của Kho dữ liệu

### **1.3. KIẾN TRÚC KHO DỮ LIỆU - DATA WAREHOUSE**

Thiết kế của Kho dữ liệu được gọi là kiến trúc Kho dữ liệu. Tùy vào nhu cầu của việc sử dụng Kho dữ liệu mà kiến trúc có thể có nhiều cấp độ khác nhau. Thông thường có các thiết kế kiến trúc cấp một, cấp hai và cấp ba [3].

Kiến trúc một tầng của Kho dữ liệu thường được sử dụng trong xử lý hàng loạt và xử lý thời gian thực để xử lý dữ liệu hoạt động. Thiết kế một tầng chỉ có một lớp phần cứng duy nhất với mục tiêu giữ cho không gian dữ liệu ở mức tối thiểu.

Trong thiết kế kiến trúc hai tầng, quá trình phân tích được tách biệt khỏi quá trình kinh doanh của doanh nghiệp, từ đó tăng cường mức độ kiểm soát và hiệu quả [3].

Thiết kế kiến trúc ba tầng bao gồm tầng trên cùng, tầng giữa và tầng dưới. Những tầng này được gọi là lớp nguồn (source layer), lớp hòa giải (reconciled layer) và lớp kho dữ liệu (data warehouse layer). Thiết kế này phù hợp với các hệ thống sử dụng lâu dài. Khi dữ liệu thay đổi, một lớp sẽ xem xét và phân tích dữ liệu bổ sung để đảm bảo không gây ra lỗi.

## **2. TÌM HIỂU HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU SQL SERVER**

### **2.1. KHÁI QUÁT VỀ SQL SERVER**

SQL Server là một hệ quản trị cơ sở dữ liệu quan hệ (Relational Database Management System) được phát triển bởi tập đoàn Microsoft nổi tiếng. Hệ quản trị cơ sở dữ liệu này cung cấp nhiều tính năng và dịch vụ để quản lý và lưu trữ dữ liệu, cho phép người dùng lưu trữ, truy xuất và xử lý dữ liệu một cách hiệu quả và đảm bảo tính an toàn [4].

### **2.2. ỨNG DỤNG SQL SERVER CHO DATA WAREHOUSE**

SQL Server cung cấp các tính năng lưu trữ dữ liệu của Data Warehouse. Các bảng và chỉ mục có thể được tạo và quản lý một cách hiệu quả để lưu trữ dữ liệu với kích thước lớn và đa dạng [5].

SQL Server có khả năng xử lý các tác vụ ETL (Extract - Transform - Load), cho phép người dùng trích xuất dữ liệu từ nhiều nguồn khác nhau, chuyển đổi dữ liệu để phù hợp với mô hình của Data Warehouse và cuối cùng là tải dữ liệu vào Data Warehouse.

Các công cụ và tính năng của SQL giúp đảm bảo tính toàn vẹn dữ liệu cũng như hiệu suất của Data Warehouse. Chẳng hạn như người dùng có thể thực hiện việc tạo chỉ mục, thống kê và quản lý khoảng trống dữ liệu để tối ưu hóa câu truy vấn và thời gian phản hồi.

SQL Server cung cấp các tính năng bảo mật rất hiệu quả như phân quyền người dùng (login) và mã hóa dữ liệu, từ đó giúp bảo vệ thông tin quan trọng trong Data Warehouse khỏi các mối đe dọa về bảo mật [5].

SQL Server còn có các công cụ và tính năng phục vụ cho việc phân tích dữ liệu trong Data Warehouse. Các câu truy vấn phức tạp và các thủ tục lưu trữ có thể được sử dụng để truy xuất và phân tích dữ liệu nhằm đưa ra quyết định kinh doanh cho doanh nghiệp.

Ngoài ra, SQL Server được tích hợp với các công nghệ khác của Microsoft như Power BI và Azure, giúp tạo ra một hệ sinh thái mạnh mẽ cho quá trình xây dựng và quản lý Data Warehouse.

### **3. TÌM HIỂU CÔNG CỤ SQL SERVER INTEGRATION SERVICES**

#### **3.1. KHÁI QUÁT VỀ SQL SERVER INTEGRATION SERVICES (SSIS)**

SQL Server Integration Services (SSIS) là dịch vụ tích hợp máy chủ SQL của Microsoft dùng để quản lý và thực hiện các quá thao tác tích hợp dữ liệu trong môi trường SQL Server với các hệ thống dữ liệu khác. SSIS được dùng để trích xuất dữ liệu từ nhiều nguồn khác nhau, chuyển đổi dữ liệu, tải dữ liệu vào cơ sở dữ liệu SQL Server hoặc các hệ thống lưu trữ dữ liệu khác [6].

#### **3.2. CHỨC NĂNG CỦA SSIS**

SSIS có tác dụng chính là đưa dữ liệu từ các nguồn dữ liệu khác nhau sang các đích khác. Công cụ này cung cấp nhiều giải pháp cho việc lưu trữ dữ liệu cho ETL để hỗ trợ các hoạt động tích hợp dữ liệu và quy trình làm việc.

SSIS được sử dụng phổ biến để thực hiện các công việc như sau:

- Lưu trữ dữ liệu: SSIS có chức năng hợp nhất dữ liệu và đồng nhất thông tin, chuyển đổi các thông tin được lưu trữ thành nguồn dữ liệu có giá trị đối với doanh nghiệp [6].
- Tải dữ liệu lớn: SSIS có một thành phần đích được dùng để tải hàng loạt dữ liệu từ các tệp phẳng (flat file) được lưu trữ trong cơ sở dữ liệu SQL hoặc tải dữ liệu hàng loạt vào hệ quản trị SQL Server. Công cụ này giúp phát hiện và báo lỗi nhanh chóng khi có lỗi xảy ra trong quá trình tải dữ liệu phức tạp.
- Lập chỉ mục dữ liệu hoặc quản lý lịch sử: Dịch vụ SSIS cho phép người dùng tạo và đặt cấu hình tùy ý cho các tác vụ chuyển đổi dữ liệu như thêm hoặc sửa bản ghi, thêm bảng, cột và hàng mới.
- Làm sạch dữ liệu: Để chuẩn hóa tất cả thông tin, ta có thể sử dụng SSIS để thực hiện các tác vụ chuyển đổi dữ liệu như làm sạch, chuyển đổi và làm phong phú. Ta cũng có thể xác định các bản ghi trùng lặp bằng tính năng chuyển đổi nhóm SSIS để xóa các bản ghi đó trước khi tải dữ liệu [6].



### 3.3. THÀNH PHẦN CHÍNH CỦA SSIS

SQL Server Integration Services (SSIS) bao gồm nhiều thành phần chính để thiết kế và thực thi các quá trình di chuyển dữ liệu. Các thành phần chính của SSIS bao gồm:

Gói SSIS (SSIS Package): Đây là phần cơ bản và quan trọng nhất của SSIS. Là nơi người dùng định nghĩa, thiết kế và tổ chức các quá trình tích hợp dữ liệu. Một gói SSIS bao gồm nhiều tác vụ và kết nối dữ liệu [6].

Tác vụ (Tasks): Là các hoạt động cụ thể trong một gói SSIS, như các tác vụ để thực hiện trích xuất dữ liệu, chuyển đổi dữ liệu, tải dữ liệu và quản lý lịch trình [6].

Luồng Dữ liệu (Data Flow): Là tác vụ cho phép người dùng trích xuất, biến đổi và nạp dữ liệu. Người có thể sử dụng các thành phần như nguồn dữ liệu, bộ biến đổi và đích dữ liệu để xây dựng luồng dữ liệu phức tạp [6].

Biến (Variables): Là các giá trị có thể được dùng để lưu trữ thông tin tạm thời trong quá trình thực thi gói SSIS. Biến có thể được dùng để truyền dữ liệu giữa các tác vụ hoặc tác vụ và luồng dữ liệu. [6]

Kết nối Dữ liệu (Connection Managers): Nó đại diện cho các thông tin kết nối tới nguồn dữ liệu hoặc đích dữ liệu. SSIS hỗ trợ nhiều loại kết nối dữ liệu như hệ quản trị cơ sở dữ liệu SQL Server, Microsoft Excel, tệp văn bản,...

Lịch trình (Schedules): Cho phép người dùng lập kế hoạch để thực thi gói SSIS. Lịch trình giúp tự động hóa quá trình tích hợp dữ liệu [6].

Quản lý gói (Package Deployment and Management): SSIS cung cấp các công cụ quản lý để triển khai, theo dõi và quản lý các gói tích hợp dữ liệu [6].

## **4. TÌM HIỂU PHẦN MỀM POWER BI**

### **4.1. KHÁI QUÁT VỀ POWER BI**

Power BI là một công cụ phân tích dữ liệu mạnh mẽ và linh hoạt do Microsoft phát triển. Nó cho phép người dùng kết nối và tổ chức dữ liệu từ nhiều nguồn khác nhau để tạo ra các báo cáo và trực quan hóa dữ liệu để hiểu rõ hơn về hoạt động kinh doanh và ra quyết định thông minh [7].

### **4.2. ỨNG DỤNG POWER BI CHO VIỆC TRỰC QUAN HÓA DỮ LIỆU**

Power BI cho phép người dùng kết nối và trích xuất dữ liệu từ Data Warehouse, sau đó tạo ra các báo cáo và biểu đồ phân tích để hiển thị thông tin quan trọng về hoạt động kinh doanh. Người dùng có thể tạo các biểu đồ cột, đường, tròn, bản đồ và nhiều loại biểu đồ khác để trực quan hóa dữ liệu từ Data Warehouse.

Power BI không chỉ kết nối với Data Warehouse, mà còn có thể kết nối với nhiều nguồn dữ liệu khác nhau như các cơ sở dữ liệu, tệp văn bản, dịch vụ web và dịch vụ điện toán đám mây. Bằng cách tích hợp các nguồn dữ liệu này, người dùng có thể tạo ra các báo cáo phân tích toàn diện hơn, kết hợp thông tin từ nhiều nguồn khác nhau [7].

Power BI cho phép người dùng tổ chức và hiển thị các báo cáo và biểu đồ trong các bảng điều khiển (Dashboards). Bằng cách tạo bảng điều khiển, người dùng có thể hiển thị các chỉ số quan trọng và theo dõi hiệu suất kinh doanh trực quan và dễ dàng.

Power BI cung cấp các tính năng tương tác mạnh mẽ, cho phép người dùng tương tác với dữ liệu và thực hiện các phân tích động. Ví dụ, người dùng có thể sử dụng các bộ lọc (slicers), biểu đồ phụ thuộc (drill-down), và các bộ lọc động để khám phá thông tin chi tiết hơn từ Data Warehouse.

## **CHƯƠNG 2. THỰC NGHIỆM XÂY DỰNG DATA WAREHOUSE**

### **1. QUY TRÌNH TIẾN HÀNH THỰC NGHIỆM**

Quy trình thực nghiệm xây dựng Data Warehouse của đề tài mô phỏng lại các giai đoạn từ thu thập dữ liệu nguồn đến xử lý, lưu trữ và trực quan hóa phục vụ cho việc phân tích, thống kê. Các bước thực hiện cụ thể như sau:

- Thu thập dữ liệu nguồn: Thiết kế cơ sở dữ liệu với các bảng chứa các thuộc tính cần thiết. Thiết lập mối quan hệ giữa các bảng dữ liệu. Đề tài gọi cơ sở dữ liệu này là cơ sở dữ liệu gốc.
- Xây dựng vùng Staging: Thiết kế cơ sở dữ liệu Staging có các bảng giống với cơ sở dữ liệu gốc, thực hiện sao chép dữ liệu từ cơ sở dữ liệu gốc vào vùng Staging.
- Xây dựng Data Warehouse: Thiết kế cơ sở dữ liệu Data Warehouse theo lược đồ hình sao gồm các bảng Dimension và bảng Fact.
- Thực hiện ETL dữ liệu vào Data Warehouse: Trích xuất, làm sạch dữ liệu Staging và tải vào Data Warehouse.
- Trực quan hóa dữ liệu: Sử dụng phần mềm Power BI để tạo báo cáo từ dữ liệu trong Data Warehouse.

### **2. THIẾT KẾ CƠ SỞ DỮ LIỆU GỐC**

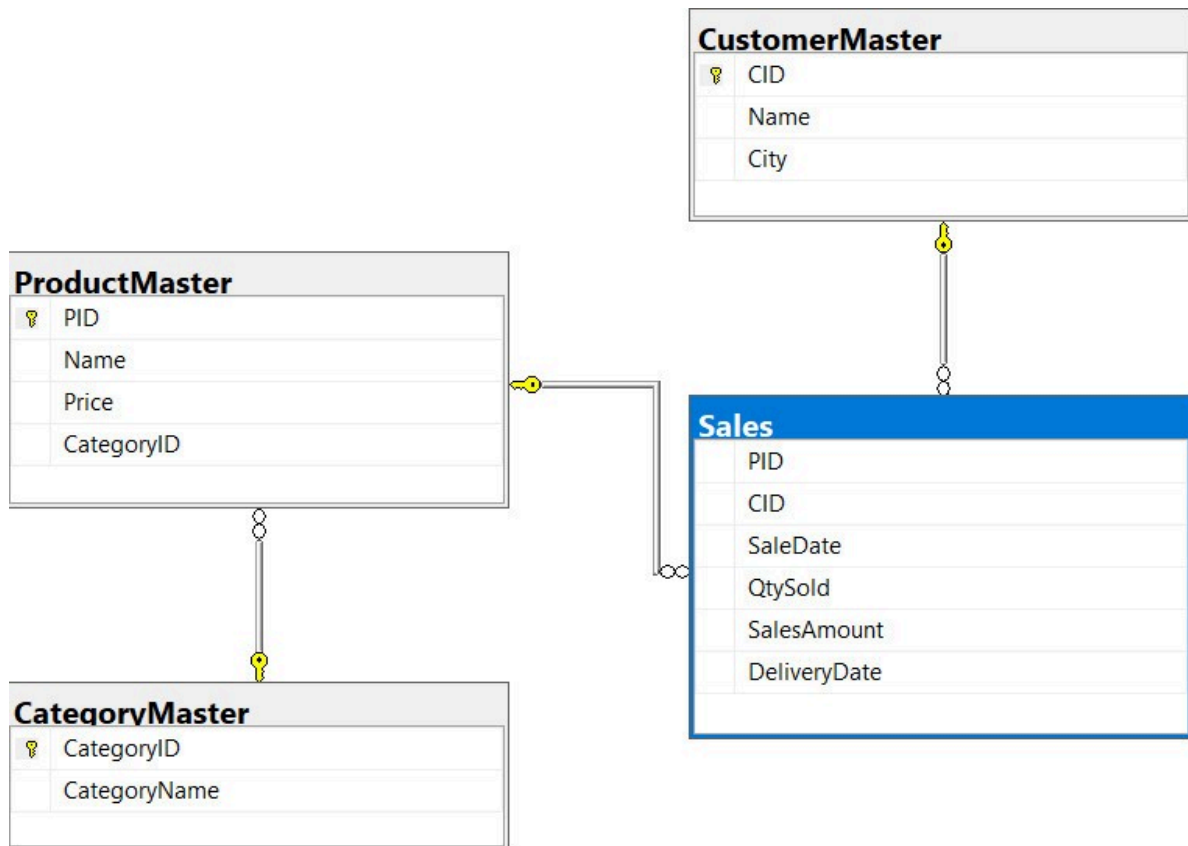
Cơ sở dữ liệu gốc có tên Beverage chứa dữ liệu từ các hoạt động thường ngày của doanh nghiệp như bán hàng, danh mục sản phẩm, danh sách sản phẩm,... chứa các bảng dữ liệu sau:

- Bảng ProductMaster với các trường:
  - PID: Mã sản phẩm
  - Name: Tên sản phẩm
  - Price: Giá bán
- Bảng CategoryMaster với các trường:
  - CategoryID: Mã loại
  - Name: Tên loại sản phẩm
- Bảng CustomerMaster với các trường:
  - CID: Mã khách hàng
  - Name: Tên khách hàng
  - City: Địa chỉ khách hàng

- Bảng Sales với các trường:
  - PID: Mã sản phẩm
  - CID: Mã khách hàng
  - SaleDate: Ngày bán
  - QtySold: Số lượng bán
  - SalesAmount: Doanh thu
  - DeliveryDate: Ngày giao hàng

Thiết lập mối quan hệ giữa bảng CategoryMaster và bảng ProductMaster sao cho một loại sản phẩm có nhiều sản phẩm, một sản phẩm chỉ thuộc về một loại sản phẩm.

Thiết lập mối quan hệ giữa bảng CustomerMaster với bảng ProductMaster và bảng Sales sao cho một khách hàng mua nhiều sản phẩm, một sản phẩm có thể được mua bởi nhiều khách hàng. Từ đó sẽ sinh ra một bảng Sales.



Hình 2.2.1. Sơ đồ quan hệ của cơ sở dữ liệu Beverage

Các bảng của cơ sở dữ liệu Beverage chứa dữ liệu như sau:

```

1  /***** Script for SelectTopNRows command from SSMS *****/
2  SELECT TOP (1000) [CategoryID]
3      , [CategoryName]
4  FROM [Beverage].[dbo].[CategoryMaster]

```

107 %

Results Messages

	CategoryID	CategoryName
1	101	Soft drink
2	102	Beer
3	103	Pure water
4	104	Others

Hình 2.2.2. Dữ liệu của bảng CategoryMaster

```

1  /***** Script for SelectTopNRows command from SSMS *****/
2  SELECT TOP (1000) [CID]
3      , [Name]
4      , [City]
5  FROM [Beverage].[dbo].[CustomerMaster]

```

107 %

Results Messages

	CID	Name	City
1	1	Praveen T	Bangalore
2	2	John Kennedy	Dallas
3	3	Suneetha K	Hyderabad
4	4	Bhuvan Gite	Mysore
5	5	Prakash Sanjay	Chennai
6	6	Vimala Roy	Baltimore
7	7	Kiran Shetty	Ooty
8	8	Jyothi M	Chatham
9	9	Amar Kutti	Dayton
10	10	Mohit Syed	Newark
11	11	Javed Ahmed	Towson

Hình 2.2.3. Một số dòng dữ liệu của bảng CustomerMaster

```
1 /***** Script for SelectTopNRows command from SSMS *****/
2 SELECT TOP (1000) [PID]
3     , [Name]
4     , [Price]
5     , [CategoryID]
6 FROM [Beverage].[dbo].[ProductMaster]
```

107 %

Results Messages

	PID	Name	Price	CategoryID
1	1	Coca cola	15.00	101
2	2	Mirinda sarsi	15.00	101
3	3	Fanta sarsi	10.00	101
4	4	Sarsi Chuong Duong	5.00	101
5	5	Mirinda orange	25.00	101
6	6	Fanta orange	50.00	101
7	7	Heniken green	20.00	102
8	8	Heniken crystal	50.00	102
9	9	Sai Gon silver	25.00	102
10	10	Sai Gon green	30.00	102
11	11	Beer 333	60.00	102
12	12	Tiger tradition	35.00	102
13	13	Tiger crystal	35.00	102

Hình 2.2.4. Một số dòng dữ liệu của bảng ProductMaster

```

1  /***** Script for SelectTopNRows command from SSMS *****/
2  SELECT TOP (1000) [PID]
3      , [CID]
4      , [SaleDate]
5      , [QtySold]
6      , [SalesAmount]
7      , [DeliveryDate]
8  FROM [Beverage].[dbo].[Sales]

```

107 %

Results Messages

	PID	CID	SaleDate	QtySold	SalesAmount	DeliveryDate
1	10	10	2019-01-03 00:00:00.000	30	6000.00	2019-01-06 00:00:00.000
2	15	12	2019-01-04 00:00:00.000	300	24000.00	2019-01-07 00:00:00.000
3	17	22	2019-01-15 00:00:00.000	30	1800.00	2019-01-18 00:00:00.000
4	21	26	2019-01-16 00:00:00.000	80	2400.00	2019-01-19 00:00:00.000
5	28	30	2019-01-17 00:00:00.000	300	6000.00	2019-01-20 00:00:00.000
6	29	36	2019-01-18 00:00:00.000	150	450000.00	2019-01-21 00:00:00.000
7	20	33	2019-01-23 00:00:00.000	30	750000.00	2019-01-26 00:00:00.000
8	11	10	2019-01-24 00:00:00.000	30	1800.00	2019-01-27 00:00:00.000
9	1	60	2019-01-25 00:00:00.000	30	18000.00	2019-01-28 00:00:00.000

Hình 2.2.5. Một số dòng dữ liệu của bảng Sales

### 3. THIẾT KẾ CƠ SỞ DỮ LIỆU STAGING

Cơ sở dữ liệu BeverageStage có các bảng dữ liệu với thiết kế giống cơ sở dữ liệu gốc Beverage nhưng không có các mối quan hệ giữa các bảng và chưa có dữ liệu. Cơ sở dữ liệu này dùng để chứa bản sao dữ liệu của cơ sở dữ liệu gốc trước khi đưa vào Data Warehouse để đảm bảo không làm ảnh hưởng đến dữ liệu gốc.

<b>StageCategoryMaster</b>	<b>StageCustomerMaster</b>
CategoryID	CID
CategoryName	Name
	City
<b>StageProductMaster</b>	<b>StageSales</b>
PID	PID
Name	CID
Price	SaleDate
CategoryID	QtySold
	SalesAmount
	DeliveryDate

Hình 2.3.1. Thiết kế của các bảng dữ liệu của cơ sở dữ liệu BeverageStage

### 4. THIẾT KẾ DATA WAREHOUSE THEO MÔ HÌNH STAR SCHEMA

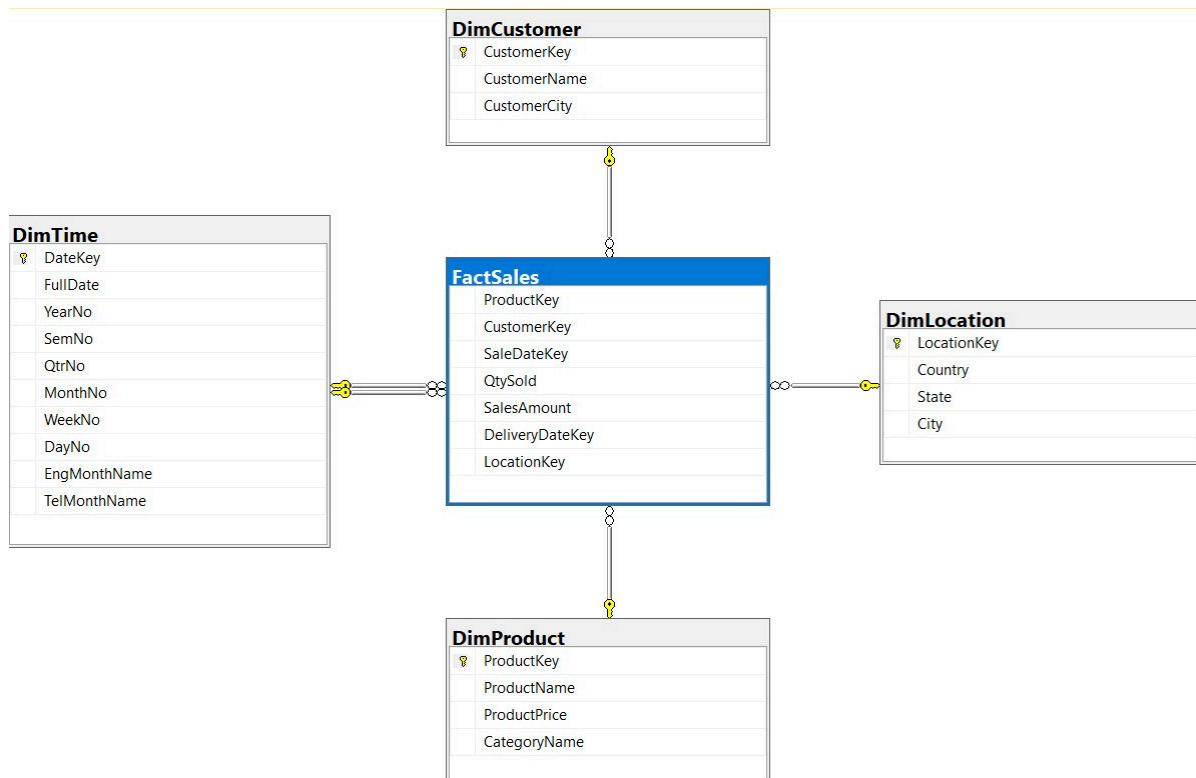
Đề tài sử dụng cơ sở dữ liệu BeverageDW làm Data Warehouse lưu trữ dữ liệu đã được xử lý từ dữ liệu thô ở cơ sở dữ liệu BeverageStage. Đồng thời, đề tài tập trung nghiên cứu xây dựng Data Warehouse theo mô hình Star Schema (lược đồ hình sao).

Các bảng Dimension của cơ sở dữ liệu BeverageDW bao gồm:

- Bảng DimTime: Lưu thông tin ngày, tháng năm, quý,... của tất cả các ngày từ năm 2019 đến năm 2025.
- Bảng DimLocation: Lưu thông tin tên quốc gia, tên tiểu bang, tên thành phố của một số quốc gia như USA, Ấn Độ,...
- Bảng DimProduct: Lưu thông tin tên sản phẩm, giá sản phẩm và tên loại sản phẩm
- Bảng DimCustomer: Lưu thông tin tên và địa chỉ khách hàng

Bảng Fact của BeverageDW là bảng FactSales lưu trữ thông tin liên quan đến hoạt động bán hàng của doanh nghiệp như số lượng bán và doanh thu.

Tất cả các bảng Dimension đều có khóa chính (Key) tương ứng. Bảng FactSales không có khóa chính. Các trường ProductKey, CustomerKey, SaleDateKey, DeliveryDateKey, LocationKey được tham chiếu từ các bảng Dimension trên, cụ thể được thể hiện thông qua sơ đồ bên dưới.



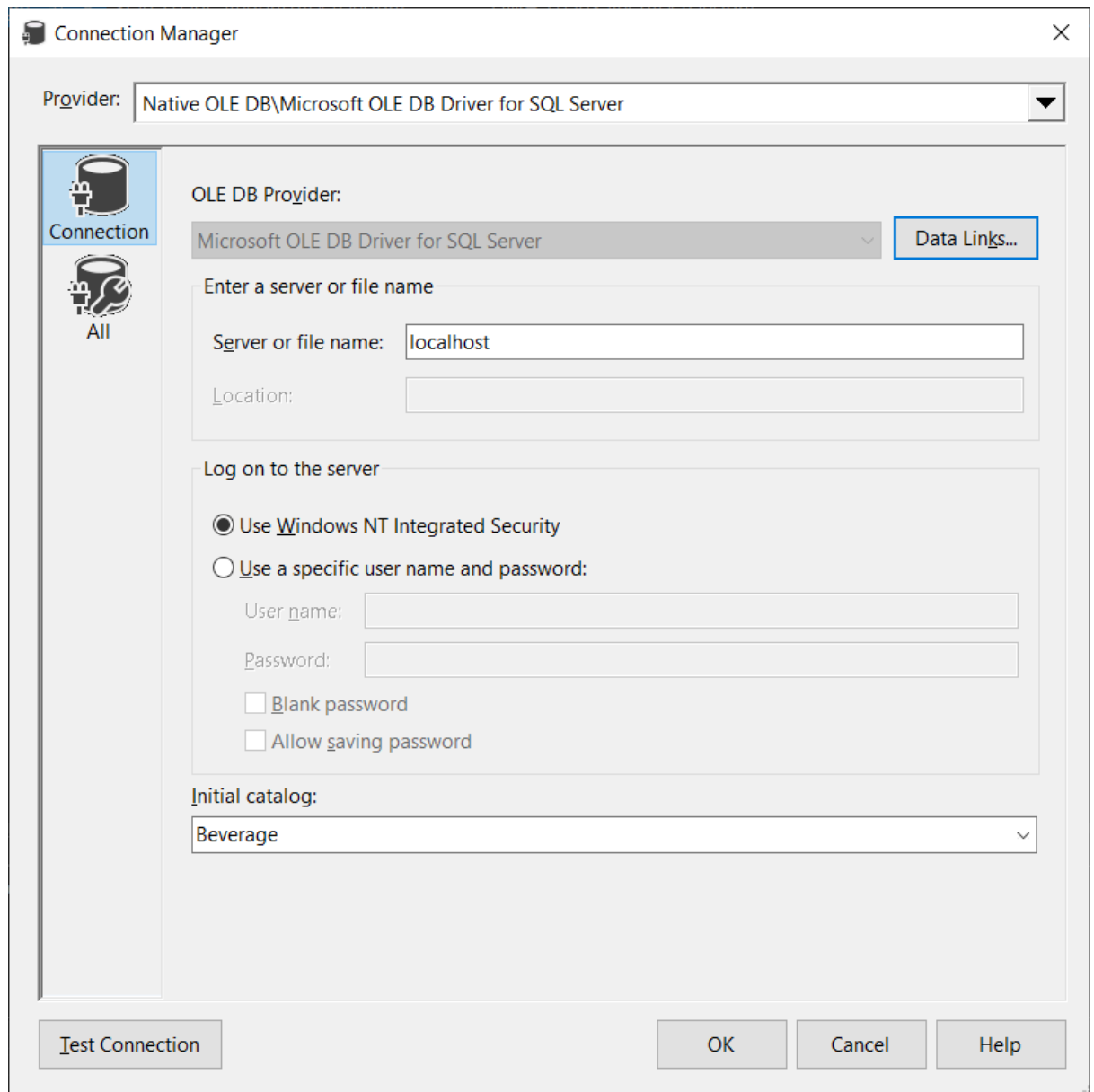
Hình 2.4.1. Sơ đồ quan hệ cơ sở dữ liệu BeverageDW

## 5. THỰC HIỆN EXTRACT - TRANSFORM - LOAD (ETL) BẰNG SQL SERVER INTEGRATION SERVICES (SSIS)

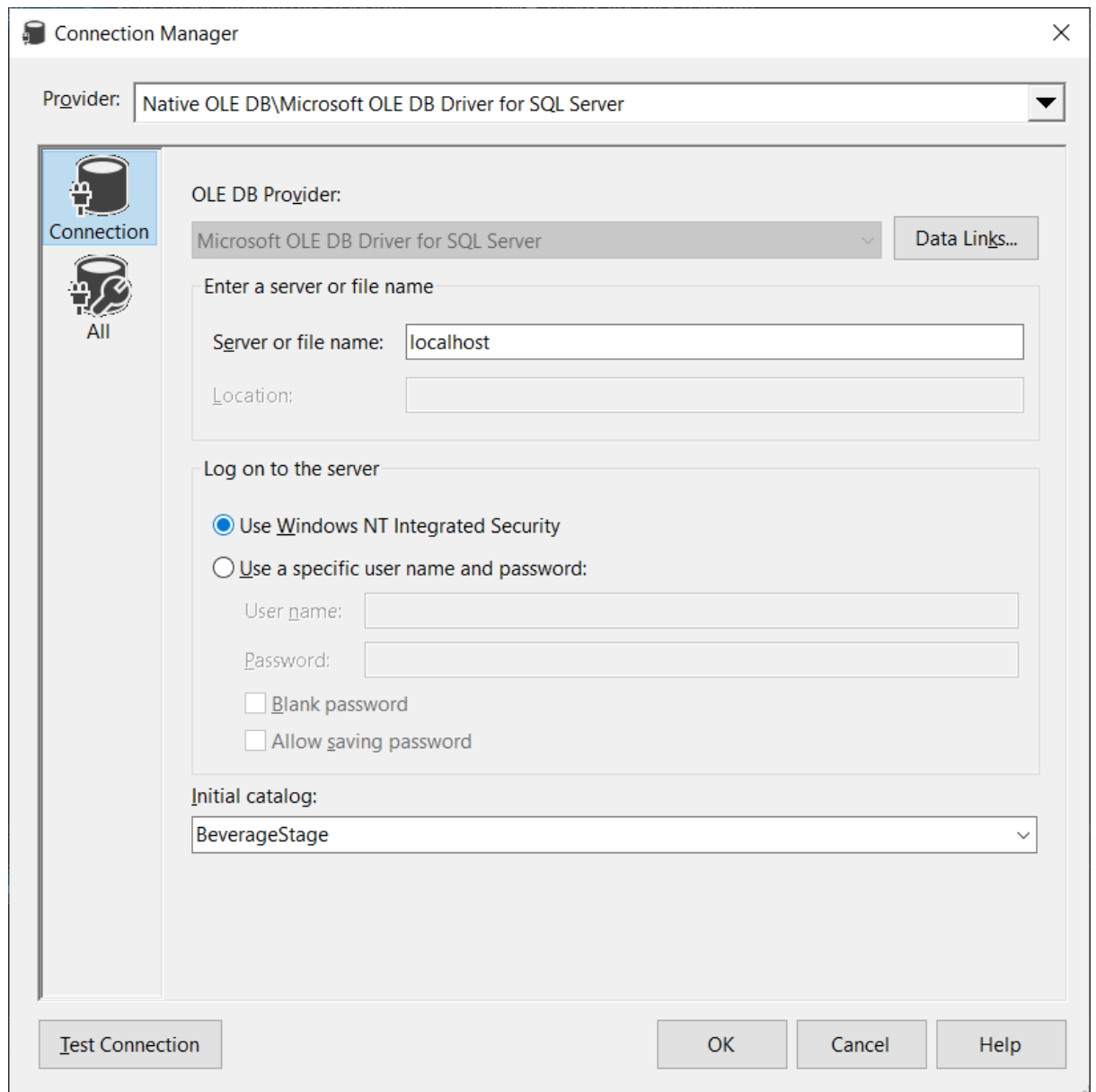
### 5.1. CẤU HÌNH CONNECTION MANAGER

Tạo kết nối từ các cơ sở dữ liệu Beverage, BeverageStage, BeverageDW lưu trữ trong SQL Server đến công cụ SSIS tại phần “Connection Manager”.

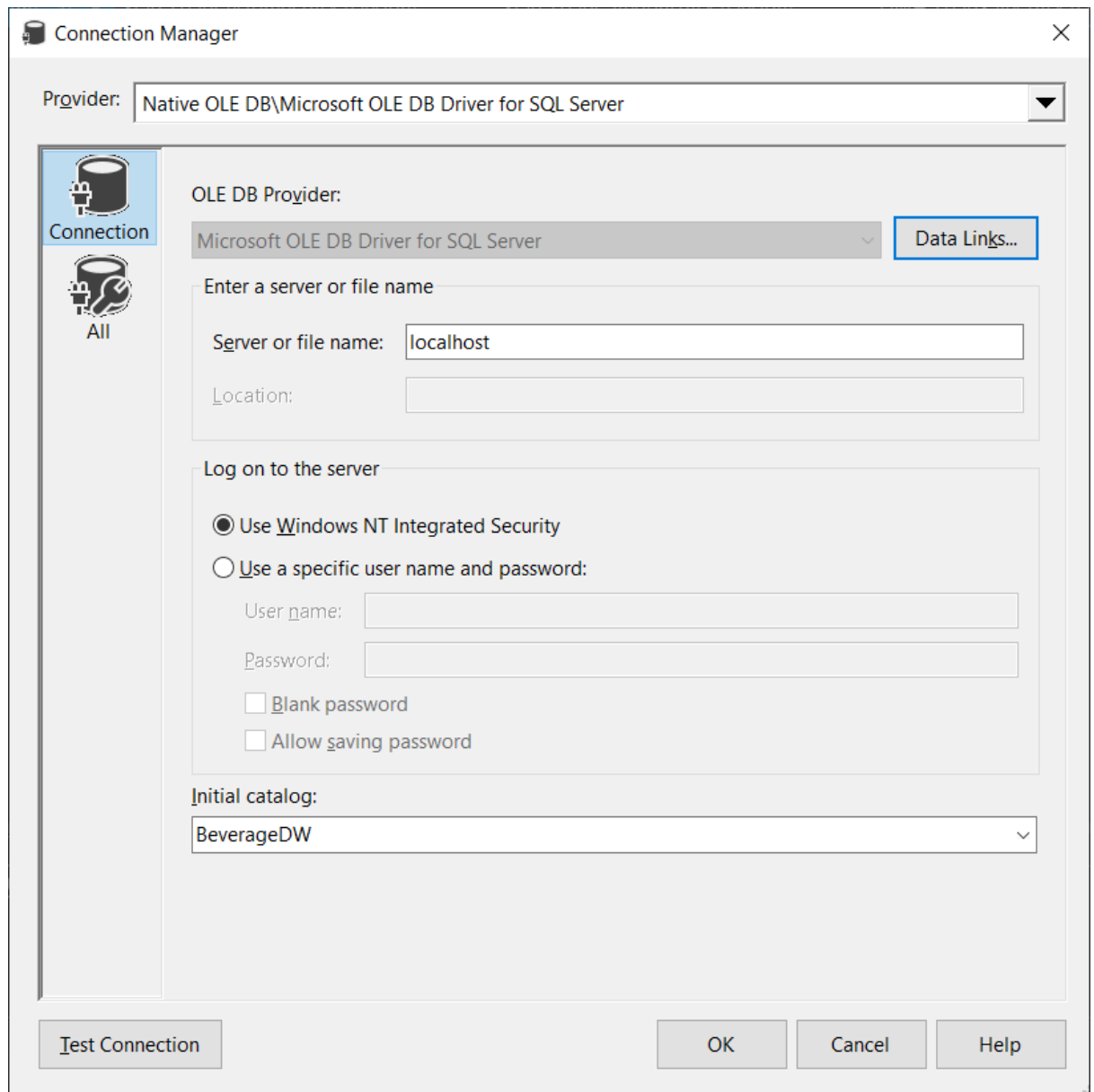




*Hình 2.5.1.1. Tạo kết nối đến cơ sở dữ liệu Beverage*



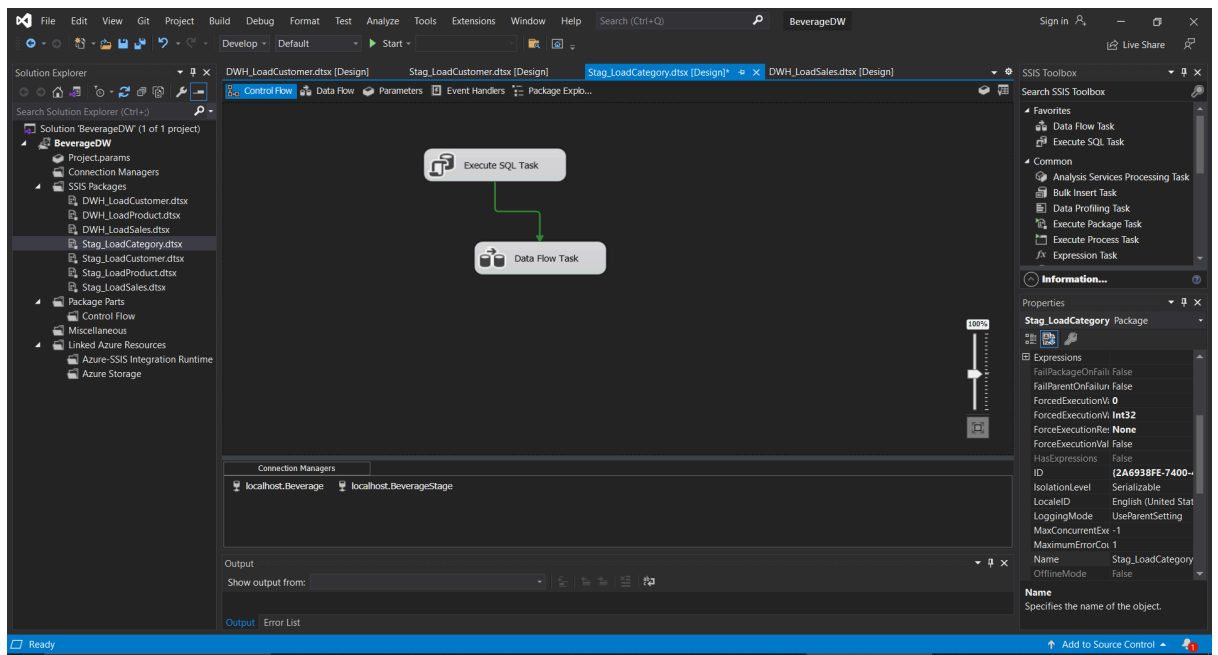
*Hình 2.5.1.2. Tạo kết nối đến cơ sở dữ liệu BeverageStage*



Hình 2.5.1.3. Tạo kết nối đến cơ sở dữ liệu BeverageDW

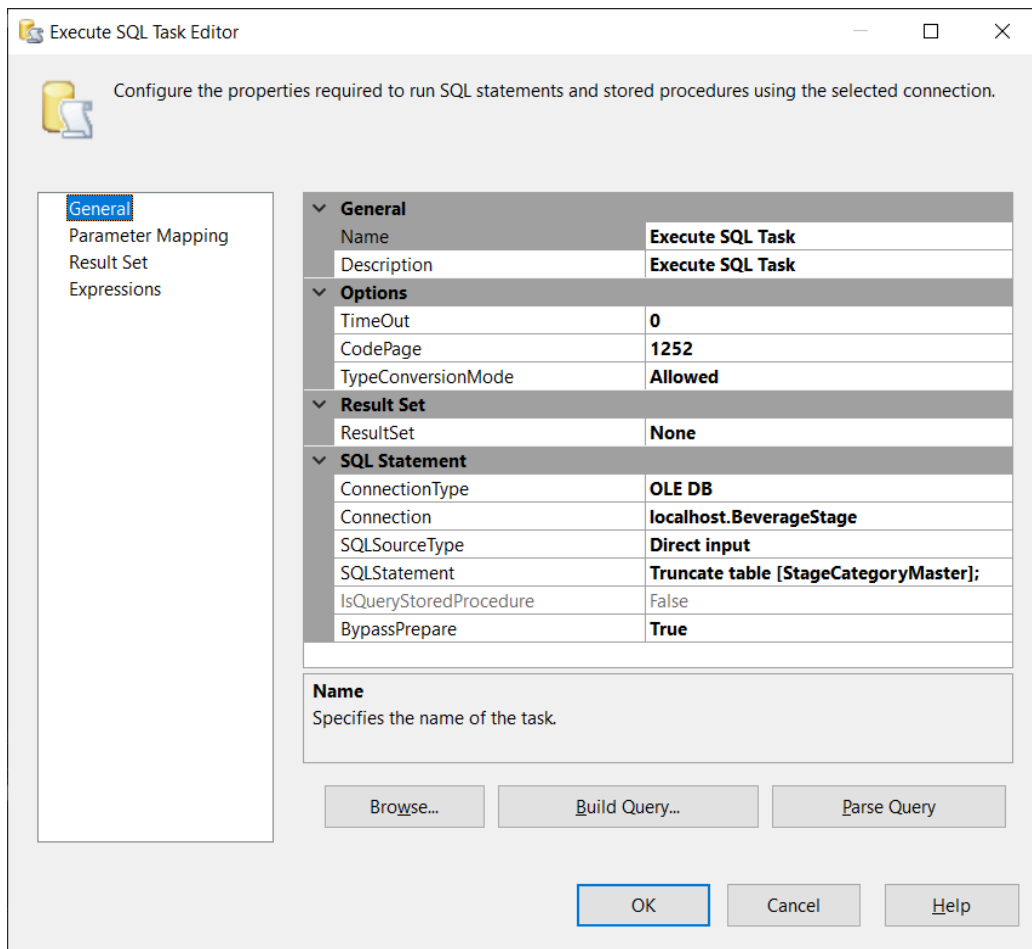
## 5.2. THỰC HIỆN SAO CHÉP DỮ LIỆU TỪ CƠ SỞ DỮ LIỆU BEVERAGE SANG BEVERAGESTAGE

Sử dụng 2 task “Execute SQL Task” và “Data Flow Task” trong phần “Control Flow” để thực hiện đổ dữ liệu tên các danh mục sản phẩm từ Beverage.CategoryMaster sang BeverageStage.StageCategoryMaster.



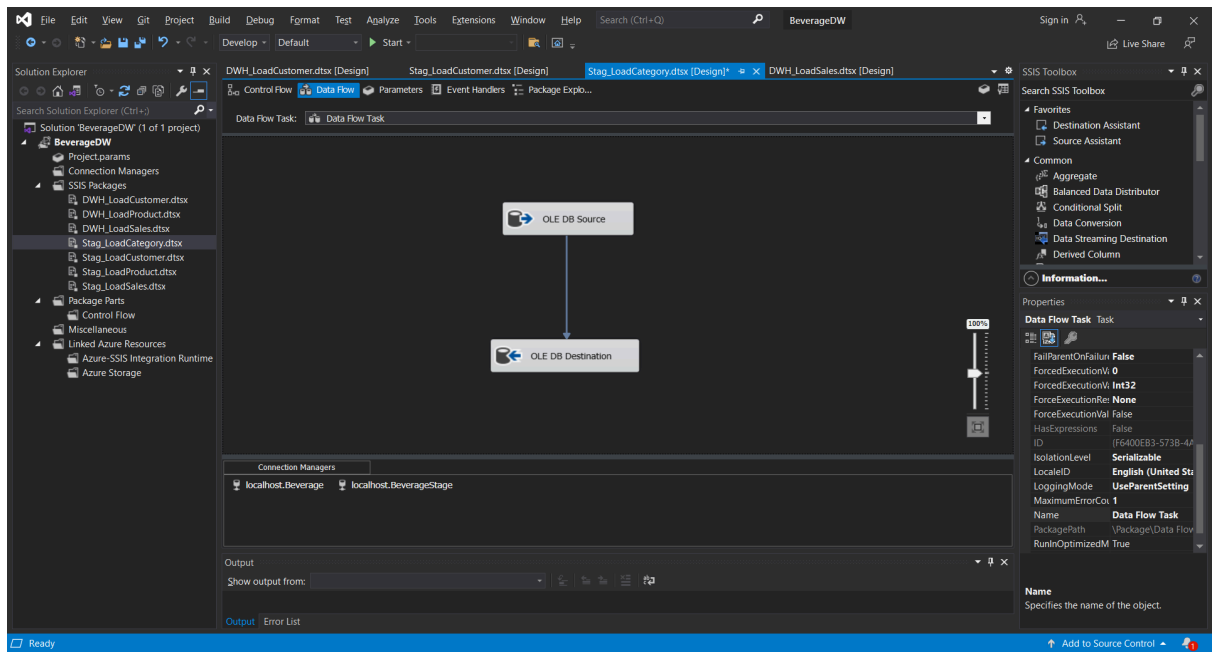
Hình 2.5.2.1. Các task trong “Control Flow”

Thêm dòng “SQL Statement” như hình bên dưới để xóa dữ liệu cũ trước khi cập nhật dữ liệu mới cho các lần cập nhật dữ liệu tiếp theo.

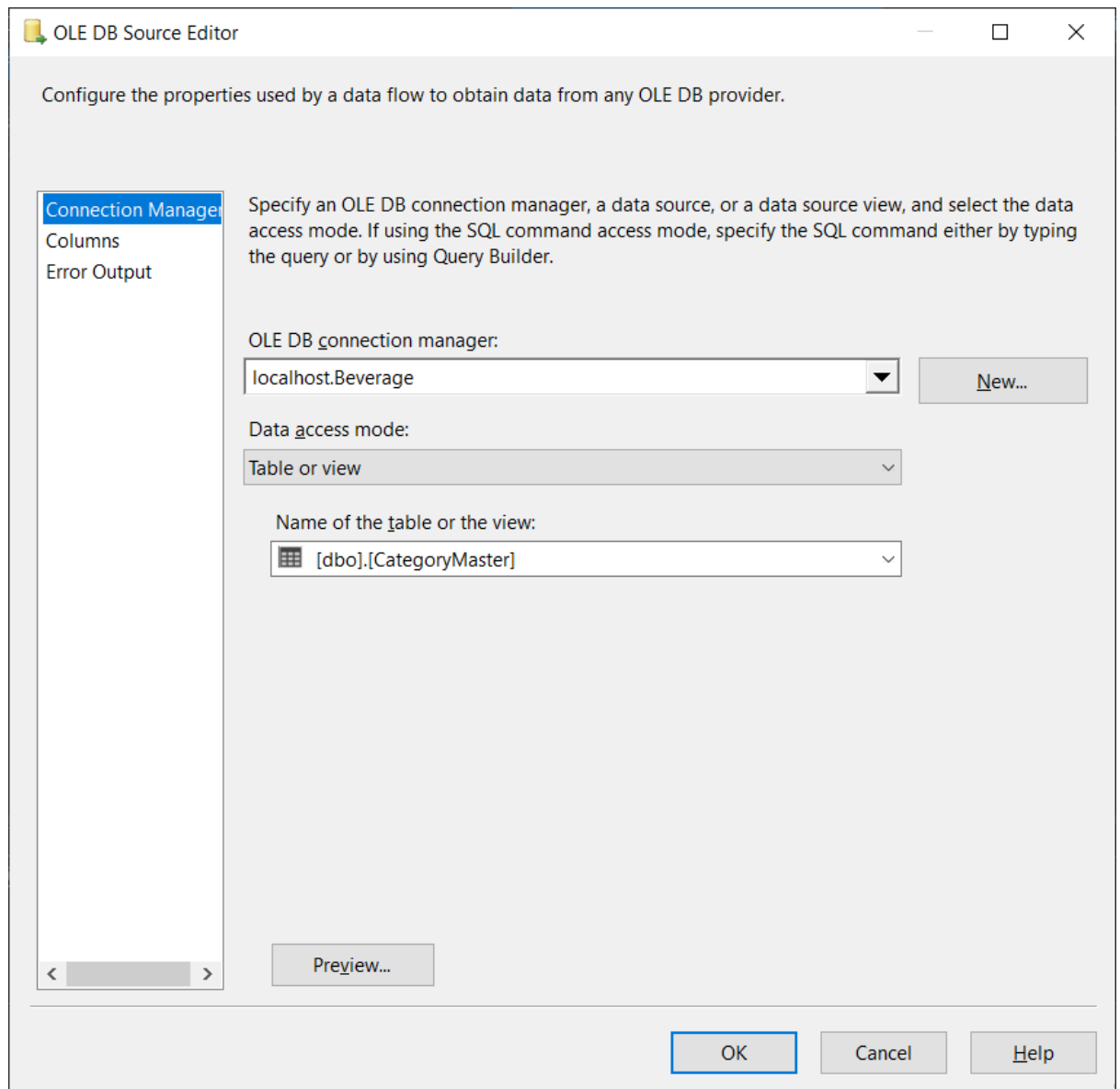


Hình 2.5.2.2. Cấu hình task “Execute SQL Task”

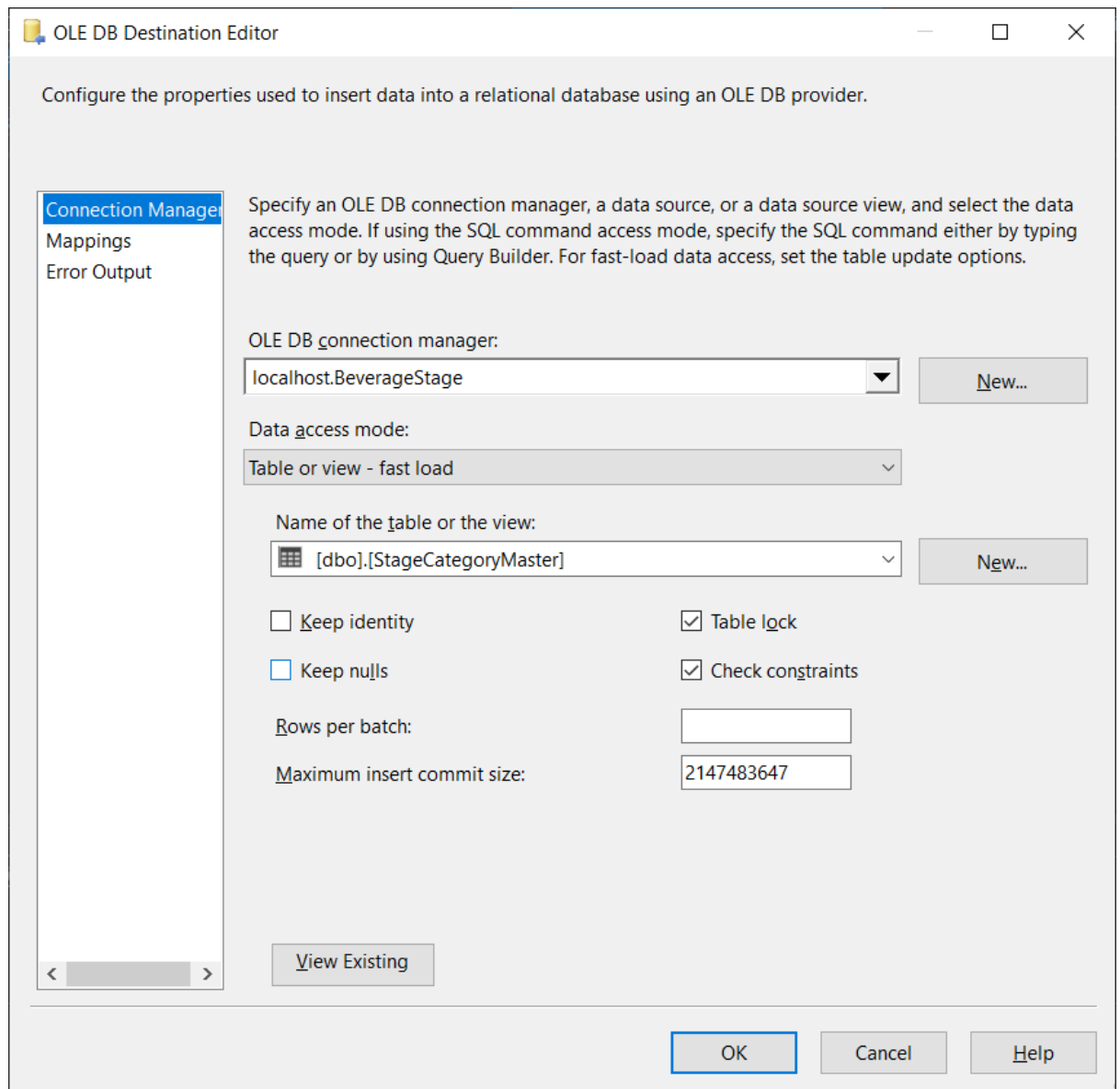
Sử dụng Toolbox “Source Assistant” và “Destination Assistant” tạo kết nối từ bảng Beverage.CategoryMaster đến BeverageStage.StageCategoryMaster.



Hình 2.5.2.3. Các Toolbox trong task “Data Flow”



*Hình 2.5.2.4. Cấu hình Toolbox “Source Assistant”*



*Hình 2.5.2.5. Cấu hình Toolbox “Destination Assistant”*

Thực hiện lặp lại các bước trên đối với các bảng StageCustomerMaster, StageProductMaster, StageSales.

Sau khi cấu hình và chạy thành công, các bảng dữ liệu ở cơ sở dữ liệu BeverageStage có dữ liệu như các hình bên dưới.

SQLQuery1.sql - LA...ALEJP7\Admin (77)

```

1 /***** Script for SelectTopNRows command from SSMS *****/
2 SELECT TOP (1000) [CategoryID]
3     , [CategoryName]
4     FROM [BeverageStage].[dbo].[StageCategoryMaster]

```

107 %

Results Messages

	CategoryID	CategoryName
1	101	Soft drink
2	102	Beer
3	103	Pure water
4	104	Others

Query executed successfully. LAPTOP-62ALEJP7 (16.0 RTM) LAPTOP-62ALEJP7\Admin ... BeverageStage 00:00:00 4 rows

Hình 2.5.2.6. Dữ liệu bảng StageCategoryMaster

SQLQuery3.sql - LA...ALEJP7\Admin (54) SQLQuery2.sql - LA...ALEJP7\Admin (53) SQLQuery1.sql - LA...ALEJP7\Admin (77)

```

1 /***** Script for SelectTopNRows command from SSMS *****/
2 SELECT TOP (1000) [CID]
3     , [Name]
4     , [City]
5     FROM [BeverageStage].[dbo].[StageCustomerMaster]

```

107 %

Results Messages

	CID	Name	City
1	1	Praveen T	Bangalore
2	2	John Kennedy	Dallas
3	3	Suneetha K	Hyderabad
4	4	Bhuvan Gite	Mysore
5	5	Prakash Sanjay	Chennai
6	6	Vimala Roy	Baltimore
7	7	Kiran Shetty	Ooty
8	8	Jyothi M	Chatham
9	9	Amar Kutli	Dayton
10	10	Mohit Syed	Newark
11	11	Javed Ahmed	Towson
12	12	Priya Raghavan	Trenton
13	13	Thulasi Ayyar	Providence
14	14	Meena Krishnan	Trenton
15	15	Malathy Srinivasan	Warwick
16	16	Yashenfal	Dayton
17	17	Akash Veri	Towson
18	18	Raghu Raghavan	Trenton
19	19	Priya Ayyar	Providence
20	20	Meenakshi khan...	Trenton
21	21	Srinivasan Raj	Warwick
22	22	Yashe Bewe	Dayton
23	23	Uvan Gitee	Mysore
24	24	Keerthi Reddy	Chennai
25	25	Prakash Raj	Chennai
26	26	Vimala Reddy	Baltimore
27	27	Shilpa Shetty	Ooty
28	28	Jyothi Makak	Chatham
29	29	Amaritha Vem	Dayton
30	30	Mohit Jain	Newark

Query executed successfully. LAPTOP-62ALEJP7 (16.0 RTM) LAPTOP-62ALEJP7\Admin ... BeverageStage 00:00:00 60 rows

Hình 2.5.2.7. Một số dòng dữ liệu bảng StageCustomerMaster



SQLQuery4.sql - LA...ALEJP7\Admin (64) SQLQuery3.sql - LA...ALEJP7\Admin (54) SQLQuery2.sql - LA...ALEJP7\Admin (53) SQLQuery1.sql - LA...ALEJP7\Admin (77)

```

1 /***** Script for SelectTopNRows command from SSMS *****/
2 SELECT TOP (1000) [PID]
3     ,[Name]
4     ,[Price]
5     ,[CategoryID]
6 FROM [BeverageStage].[dbo].[StageProductMaster]

```

107 %

Results Messages

	PID	Name	Price	CategoryID
1	1	Coca cola	15.00	101
2	2	Mirinda sarsi	15.00	101
3	3	Fanta sarsi	10.00	101
4	4	Sarsi Chuong Duong	5.00	101
5	5	Mirinda orange	25.00	101
6	6	Fanta orange	50.00	101
7	7	Heniken green	20.00	102
8	8	Heniken crystal	50.00	102
9	9	Sai Gon silver	25.00	102
10	10	Sai Gon green	30.00	102
11	11	Beer 333	60.00	102
12	12	Tiger tradition	35.00	102
13	13	Tiger crystal	35.00	102
14	14	Tiger silver	110.00	102
15	15	Tiger brown	80.00	102
16	16	Budweiser	60.00	102
17	17	Aquafina small	60.00	103
18	18	Aquafina big	200.00	103
19	19	Lavie big	300.00	103
20	20	Lavie small	50.00	103
21	21	Dasani small	30.00	103
22	22	Dasani big	150.00	103
23	23	Bird nest drink	10.00	104
24	24	Apple juice	30.00	104
25	25	Lemonade	50.00	104
26	26	Da Lat strawberry j...	100.00	104
27	27	strawberry juice	30.00	104
28	28	Pineapple juice	20.00	104
29	29	Sugar cane juice	15.00	104

Query executed successfully. LAPTOP-62ALEJP7 (16.0 RTM) LAPTOP-62ALEJP7\Admin ... BeverageStage 00:00:00 29 rows

Hình 2.5.2.8. Một số dòng dữ liệu bảng StageProductMaster

SQLQuery5.sql - LA...ALEJP7\Admin (73) SQLQuery4.sql - LA...ALEJP7\Admin (64) SQLQuery3.sql - LA...ALEJP7\Admin (54) SQLQuery2.sql - LA...ALEJP7\Admin (53)

```

1 /***** Script for SelectTopNRows command from SSMS *****/
2 SELECT TOP (1000) [PID]
3     ,[CID]
4     ,[SaleDate]
5     ,[QtySold]
6     ,[SalesAmount]
7     ,[DeliveryDate]
8 FROM [BeverageStage].[dbo].[StageSales]

```

107 %

Results Messages

	PID	CID	SaleDate	QtySold	SalesAmount	DeliveryDate
1	10	10	2019-01-03 00:00:00.000	30	6000.00	2019-01-06 00:00:00.000
2	15	12	2019-01-04 00:00:00.000	300	24000.00	2019-01-07 00:00:00.000
3	17	22	2019-01-15 00:00:00.000	30	1800.00	2019-01-18 00:00:00.000
4	21	26	2019-01-16 00:00:00.000	80	2400.00	2019-01-19 00:00:00.000
5	28	30	2019-01-17 00:00:00.000	300	6000.00	2019-01-20 00:00:00.000
6	29	36	2019-01-18 00:00:00.000	150	45000.00	2019-01-21 00:00:00.000
7	20	33	2019-01-23 00:00:00.000	30	750000.00	2019-01-26 00:00:00.000
8	11	10	2019-01-24 00:00:00.000	30	1800.00	2019-01-27 00:00:00.000
9	1	60	2019-01-25 00:00:00.000	30	18000.00	2019-01-28 00:00:00.000
10	2	10	2019-01-26 00:00:00.000	30	300.00	2019-01-29 00:00:00.000
11	4	10	2019-02-03 00:00:00.000	30	810000.00	2019-02-06 00:00:00.000
12	5	12	2019-02-04 00:00:00.000	100	10000.00	2019-02-07 00:00:00.000
13	12	22	2019-02-15 00:00:00.000	30	1050.00	2019-02-18 00:00:00.000
14	8	26	2019-02-16 00:00:00.000	80	8000.00	2019-02-19 00:00:00.000
15	9	30	2019-02-17 00:00:00.000	300	30000.00	2019-02-20 00:00:00.000
16	3	36	2019-02-18 00:00:00.000	150	1800000.00	2019-02-21 00:00:00.000
17	1	33	2019-02-23 00:00:00.000	30	750000.00	2019-02-26 00:00:00.000
18	6	10	2019-02-24 00:00:00.000	50	5000.00	2019-02-27 00:00:00.000
19	3	60	2019-02-25 00:00:00.000	100	20000.00	2019-02-28 00:00:00.000
20	7	10	2019-02-26 00:00:00.000	90	900.00	2019-03-01 00:00:00.000
21	6	10	2020-03-03 00:00:00.000	30	3000.00	2020-03-06 00:00:00.000
22	10	12	2020-03-04 00:00:00.000	300	9000.00	2020-03-07 00:00:00.000
23	10	22	2020-03-15 00:00:00.000	30	1800.00	2020-03-18 00:00:00.000
24	26	26	2020-03-16 00:00:00.000	80	8000.00	2020-03-19 00:00:00.000
25	28	30	2020-03-17 00:00:00.000	300	6000.00	2020-03-20 00:00:00.000
26	6	36	2020-03-18 00:00:00.000	150	4500.00	2020-03-21 00:00:00.000
27	8	33	2020-03-23 00:00:00.000	30	2250.00	2020-03-26 00:00:00.000

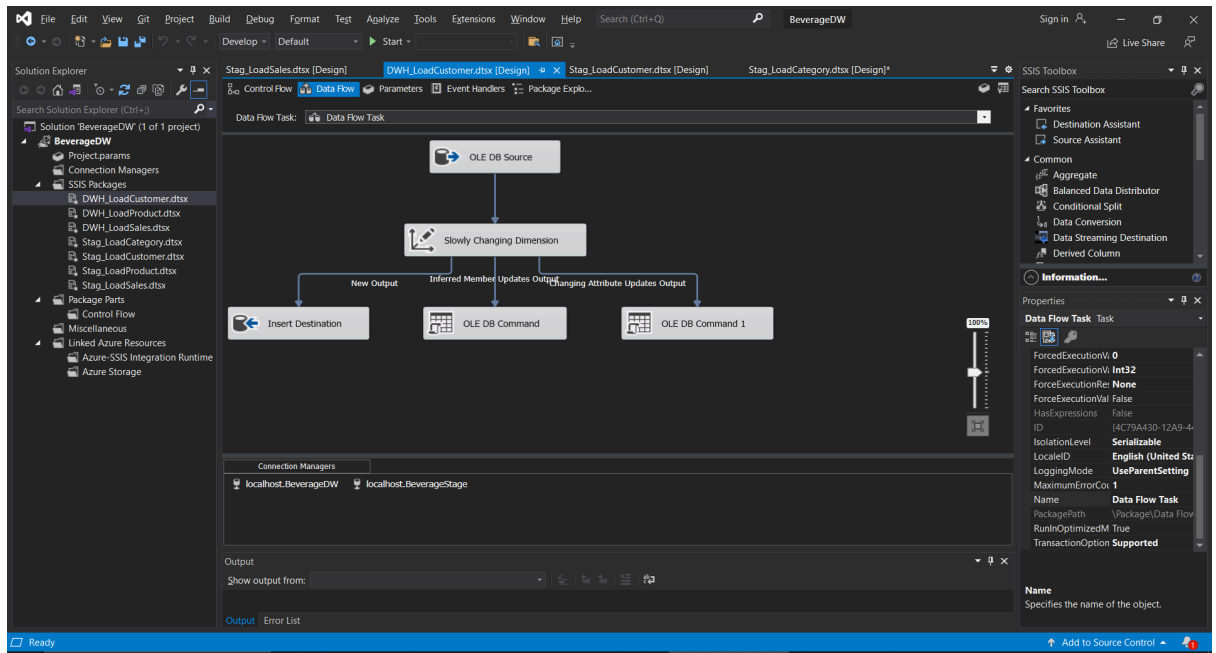
Query executed successfully. LAPTOP-62ALEJP7 (16.0 RTM) LAPTOP-62ALEJP7\Admin ... BeverageStage 00:00:00 100 rows

Hình 2.5.2.9. Một số dòng dữ liệu bảng StageSales

## 5.3. CHUYỂN ĐỔI VÀ TẢI DỮ LIỆU TỪ BEVERAGESTAGE VÀO BEVERAGEDW

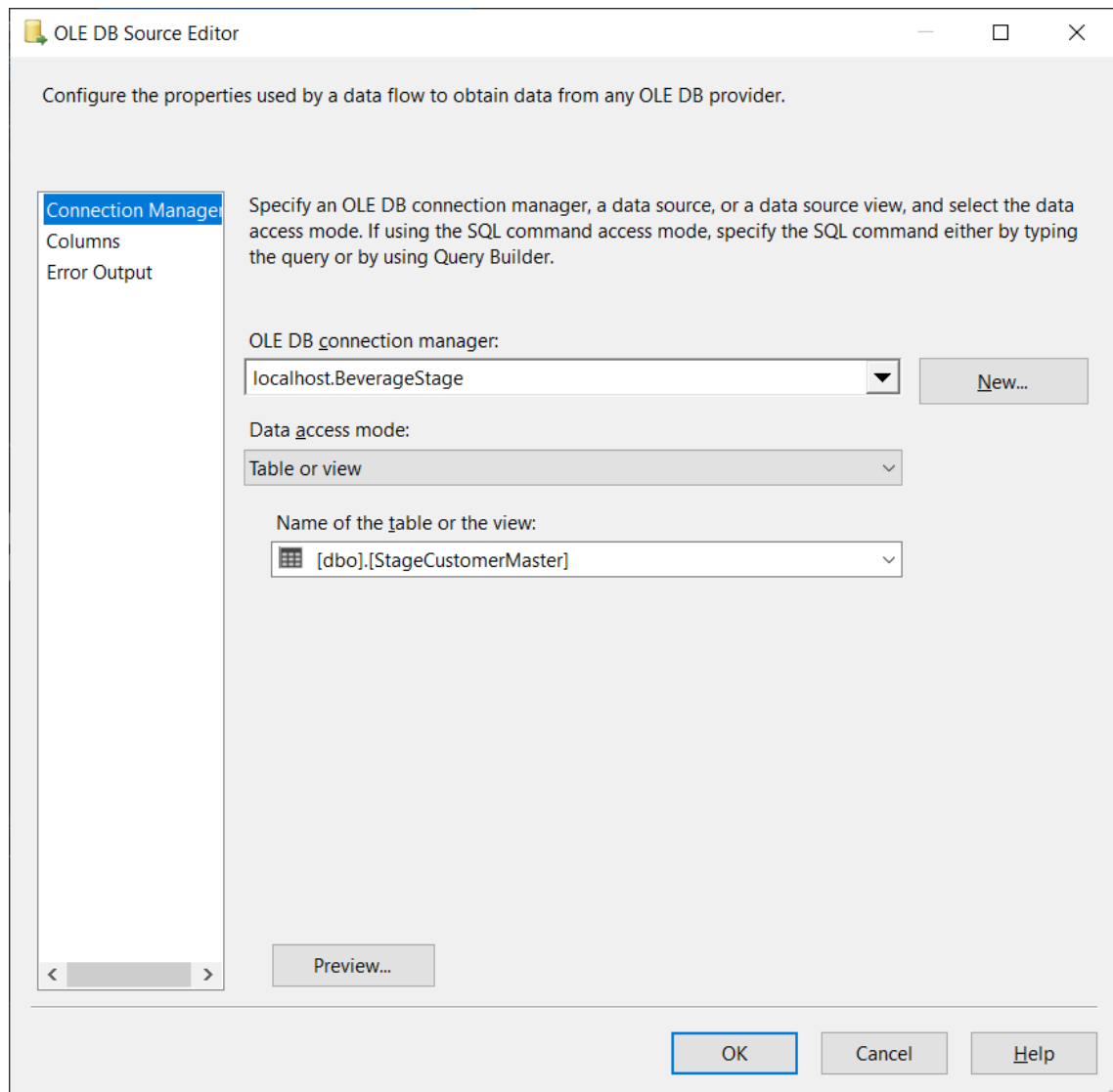
### 5.3.1. TẢI DỮ LIỆU TỪ KHÁCH HÀNG VÀO DATA WAREHOUSE

Sử dụng và cấu hình task “Data Flow” như hình bên dưới để tiến hành đổ dữ liệu từ bảng BeverageStage.StageCustomerMaster vào bảng BeverageDW.DimCustomer.



Hình 2.5.3.1.1. Cấu hình task “Data Flow”

Chỉ định bảng dữ liệu BeverageStage.StageCustomerMaster trong Toolbox “Source Assistant” và bảng BeverageDW.DimCustomer trong “Destination Assistant”. Sử dụng Toolbox “Slowly Changing Dimension” để bỏ qua dữ liệu trùng cho những lần cập nhật dữ liệu tiếp theo.



*Hình 2.5.3.1.2. Cấu hình Toolbox “Source Assistant”*

**Slowly Changing Dimension Wizard**

**Select a Dimension Table and Keys**  
Select a dimension table to load and map columns in the transformation input to

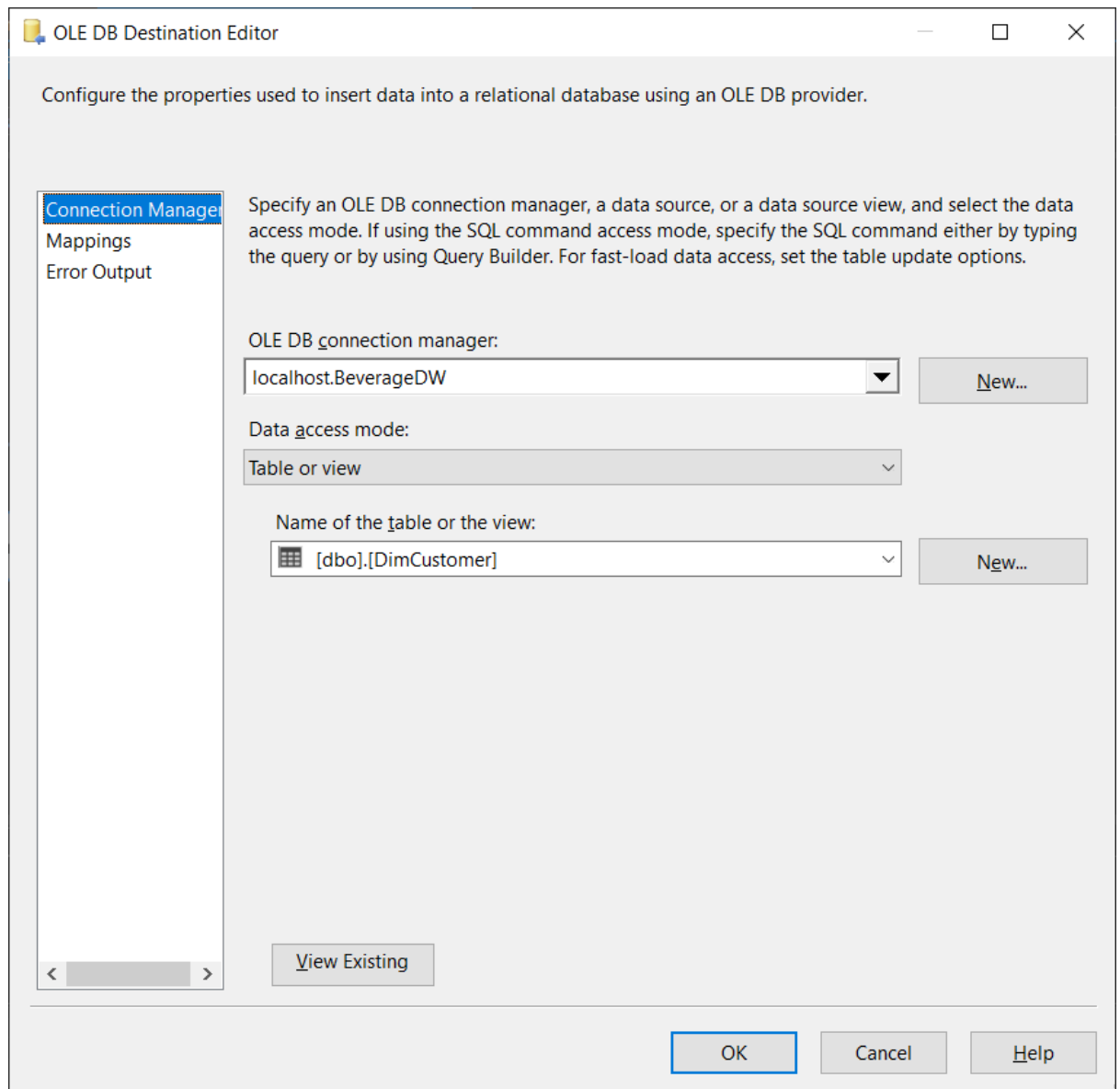
Connection manager:  
localhost.BeverageDW New...

Table or view:  
[dbo].[DimCustomer]

Input Columns	Dimension Columns	Key Type
City	CustomerCity	Not a key column
CID	CustomerKey	Business key
Name	CustomerName	Not a key column

Help < Back Next > Finish >> Cancel

Hình 2.5.3.1.2. Cấu hình Toolbox “Slowly Changing Dimension”



*Hình 2.5.3.1.3. Cấu hình Toolbox “Destination Assistant”*

Sau khi cấu hình và chạy thành công, bảng DimCustomer có dữ liệu như hình bên dưới.

1	/***** Script for SelectTopNRows command from SSMS *****/		
2	SELECT TOP (1000) [CustomerKey]		
3	, [CustomerName]		
4	, [CustomerCity]		
5	FROM [BeverageDW].[dbo].[DimCustomer]		

107 %

ResultsMessages

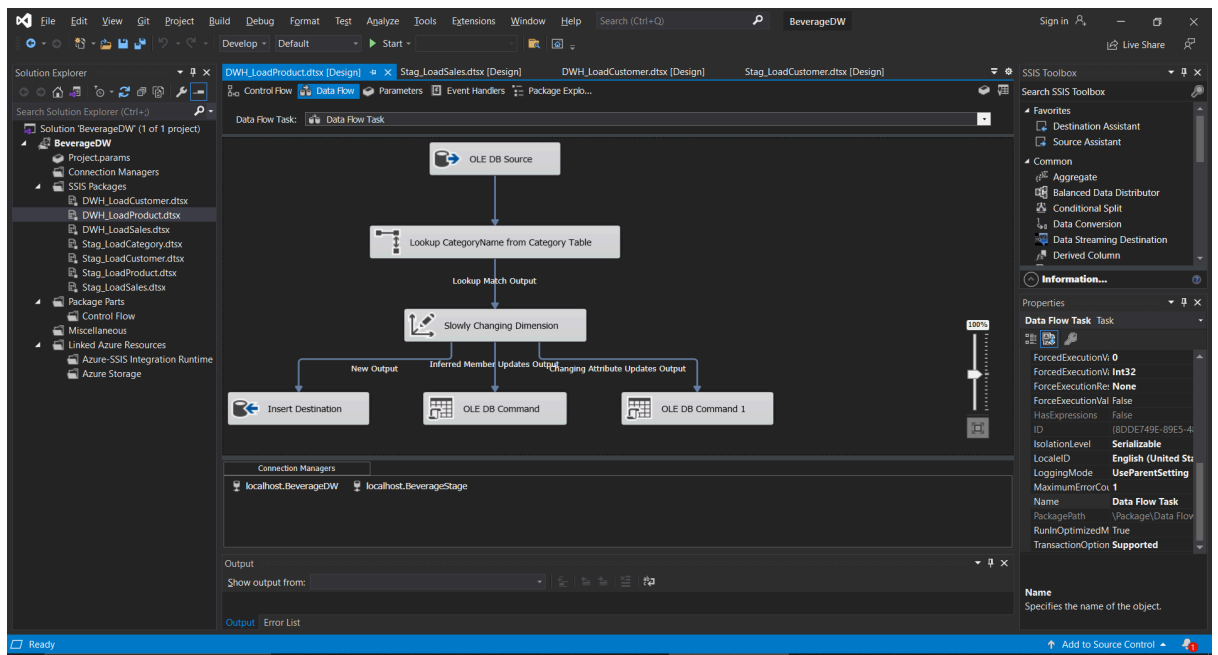
	CustomerKey	CustomerName	CustomerCity
1	1	Praveen T	Bangalore
2	2	John Kennedy	Dallas
3	3	Suneetha K	Hyderabad
4	4	Bhuvan Gite	Mysore
5	5	Prakash Sanjay	Chennai
6	6	Vimala Roy	Baltimore
7	7	Kiran Shetty	Ooty
8	8	Jyothi M	Chatham
9	9	Amar Kutti	Dayton
10	10	Mohit Syed	Newark
11	11	Javed Ahmed	Towson
12	12	Priya Raghavan	Trenton

*Hình 2.5.3.1.4. Một số dòng dữ liệu của bảng DimCustomer*

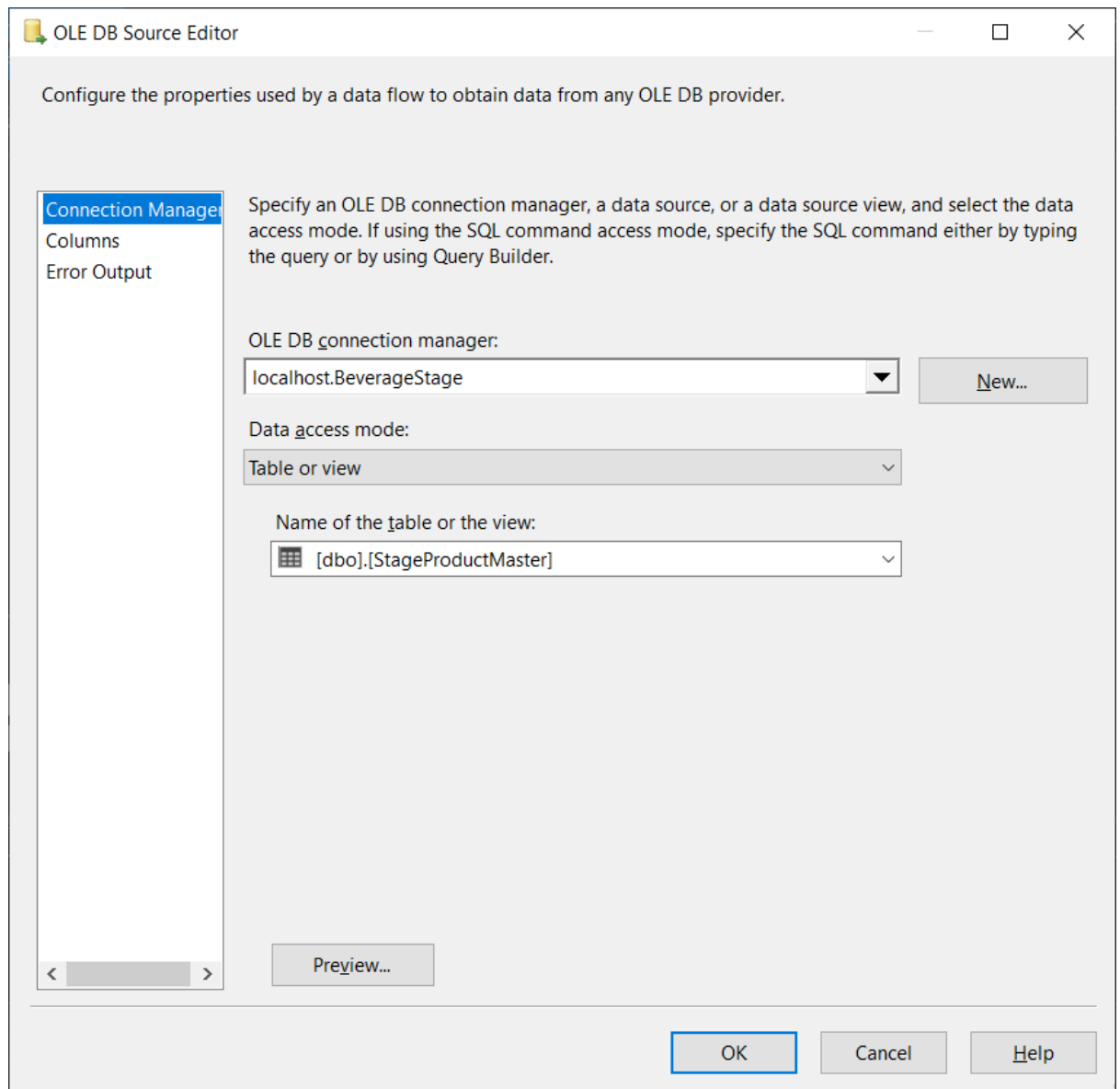
### 5.3.2. TẢI DỮ LIỆU SẢN PHẨM VÀO DATA WAREHOUSE

Quá trình tải dữ liệu từ bảng BeverageStage.StageProductMaster vào bảng BeverageDW.DimProduct cũng sử dụng các Toolbox trong “Data Flow” tương tự quá trình tải dữ liệu khách hàng.

Quá trình này có sử dụng thêm Toolbox “Lookup” tạo thêm cột mới cho DimProduct có tên “CategoryName” và lấy dữ liệu tên danh mục sản phẩm từ bảng BeverageStage.StageCategoryMaster.

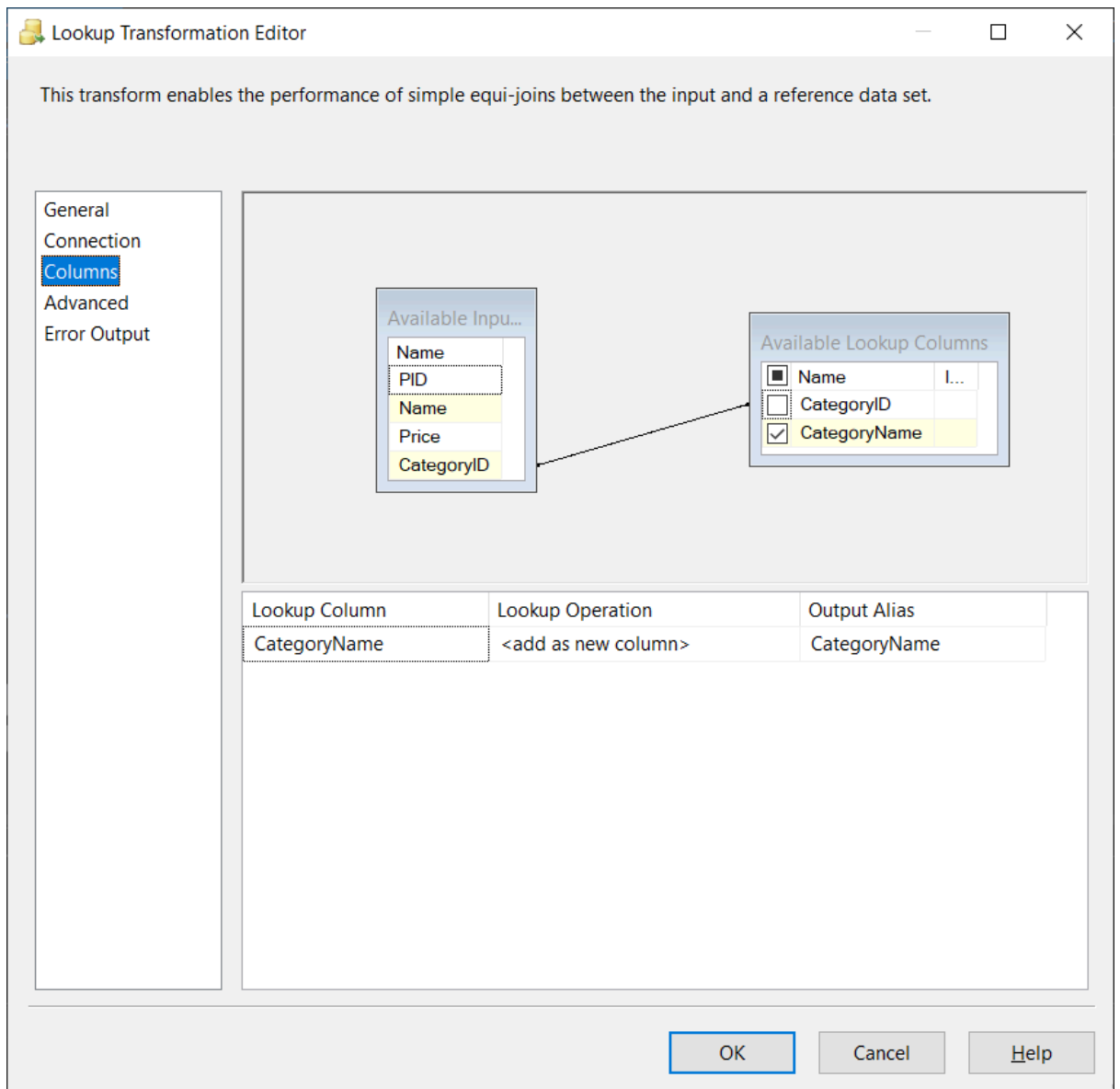


Hình 2.5.3.2.1. Cấu hình task “Data Flow”



*Hình 2.5.3.2.2. Cấu hình Toolbox “Source Assistant”*





Hình 2.5.3.2.3. Cấu hình Toolbox “Lookup”

**Slowly Changing Dimension Wizard**

**Select a Dimension Table and Keys**  
Select a dimension table to load and map columns in the transformation input to

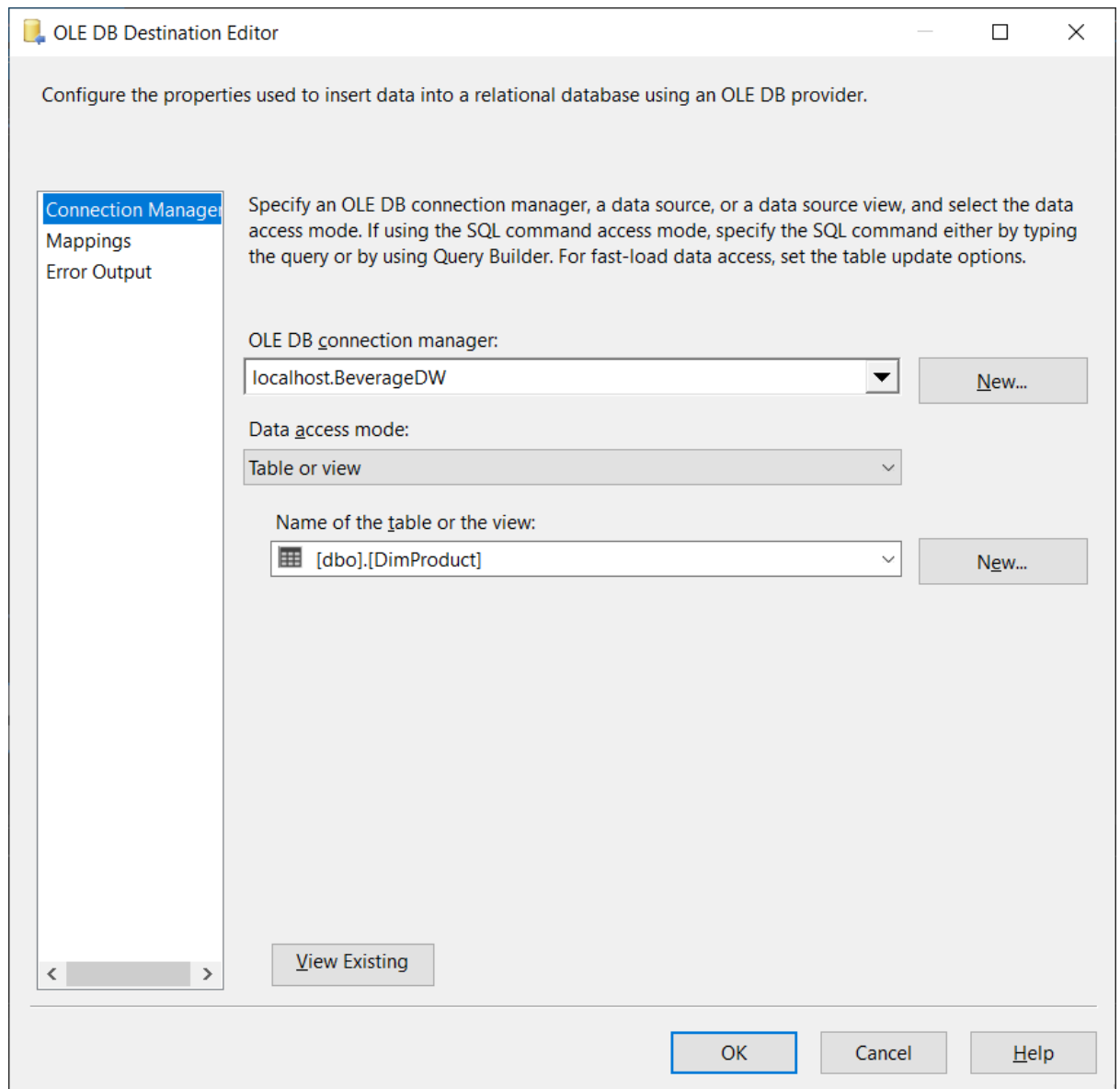
Connection manager:  
localhost.BeverageDW New...

Table or view:  
[dbo].[DimProduct]

Input Columns	Dimension Columns	Key Type
CategoryNa...	CategoryName	Not a key column
PID	ProductKey	Business key
Name	ProductName	Not a key column
Price	ProductPrice	Not a key column

Help < Back Next > Finish >> Cancel

Hình 2.5.3.2.4. Cấu hình Toolbox “Slowly Changing Dimension”



*Hình 2.5.3.2.5. Cấu hình Toolbox “Destination Assistant”*

Sau hoàn thành các cấu hình cần thiết và chạy thành công, bảng DimProduct có dữ liệu như hình bên dưới.

1	/***** Script for SelectTopNRows command from SSMS *****/			
2	SELECT TOP (1000) [ProductKey]			
3	, [ProductName]			
4	, [ProductPrice]			
5	, [CategoryName]			
6	FROM [BeverageDW].[dbo].[DimProduct]			

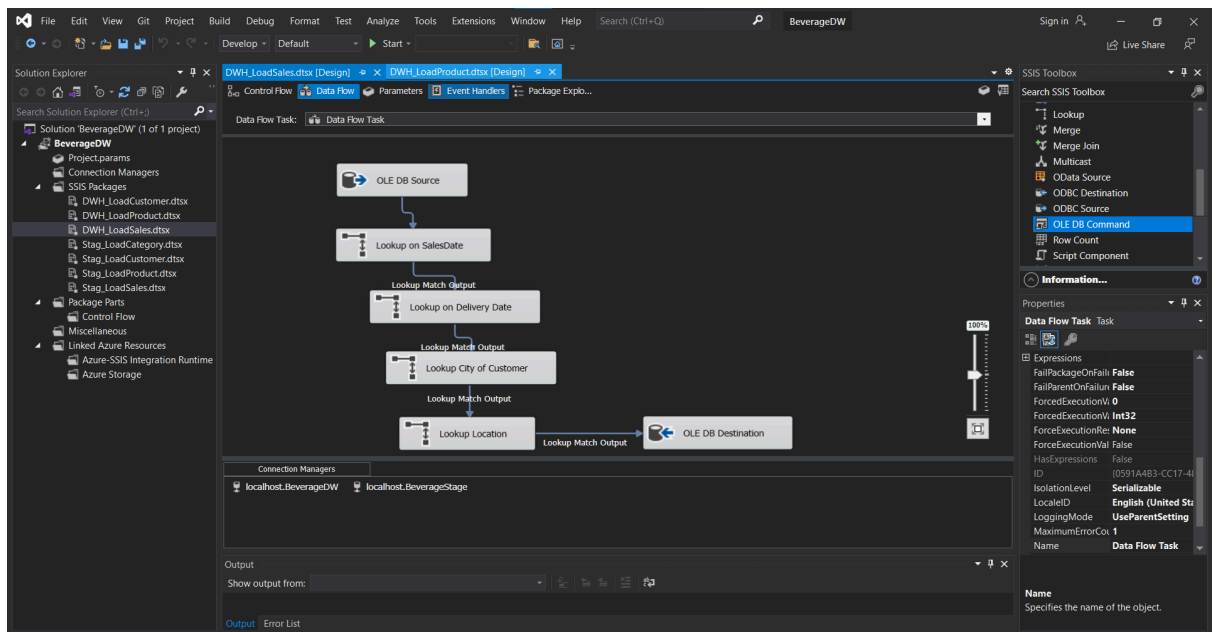
Results				
	ProductKey	ProductName	ProductPrice	CategoryName
1	1	Coca cola	15.00	Soft drink
2	2	Mirinda sarsi	15.00	Soft drink
3	3	Fanta sarsi	10.00	Soft drink
4	4	Sarsi Chuong Duong	5.00	Soft drink
5	5	Mirinda orange	25.00	Soft drink
6	6	Fanta orange	50.00	Soft drink
7	7	Heniken green	20.00	Beer
8	8	Heniken crystal	50.00	Beer
9	9	Sai Gon silver	25.00	Beer
10	10	Sai Gon green	30.00	Beer
11	11	Beer 333	60.00	Beer
12	12	Tiger tradition	35.00	Beer
13	13	Tiger crystal	35.00	Beer
14	14	Tiger silver	110.00	Beer
15	15	Tiger brown	80.00	Beer
16	16	Budweiser	60.00	Beer
17	17	Aquafina small	60.00	Pure water
18	18	Aquafina big	200.00	Pure water
19	19	Lavie big	300.00	Pure water
20	20	Lavie small	50.00	Pure water
21	21	Dasani small	30.00	Pure water
22	22	Dasani big	150.00	Pure water
23	23	Bird nest drink	10.00	Others
24	24	Apple juice	30.00	Others
25	25	Lemonade	50.00	Others
26	26	Da Lat strawberry j...	100.00	Others
27	27	strawberry juice	30.00	Others
28	28	Pineapple juice	20.00	Others

*Hình 2.5.3.2.6. Một số dòng dữ liệu của bảng DimProduct*

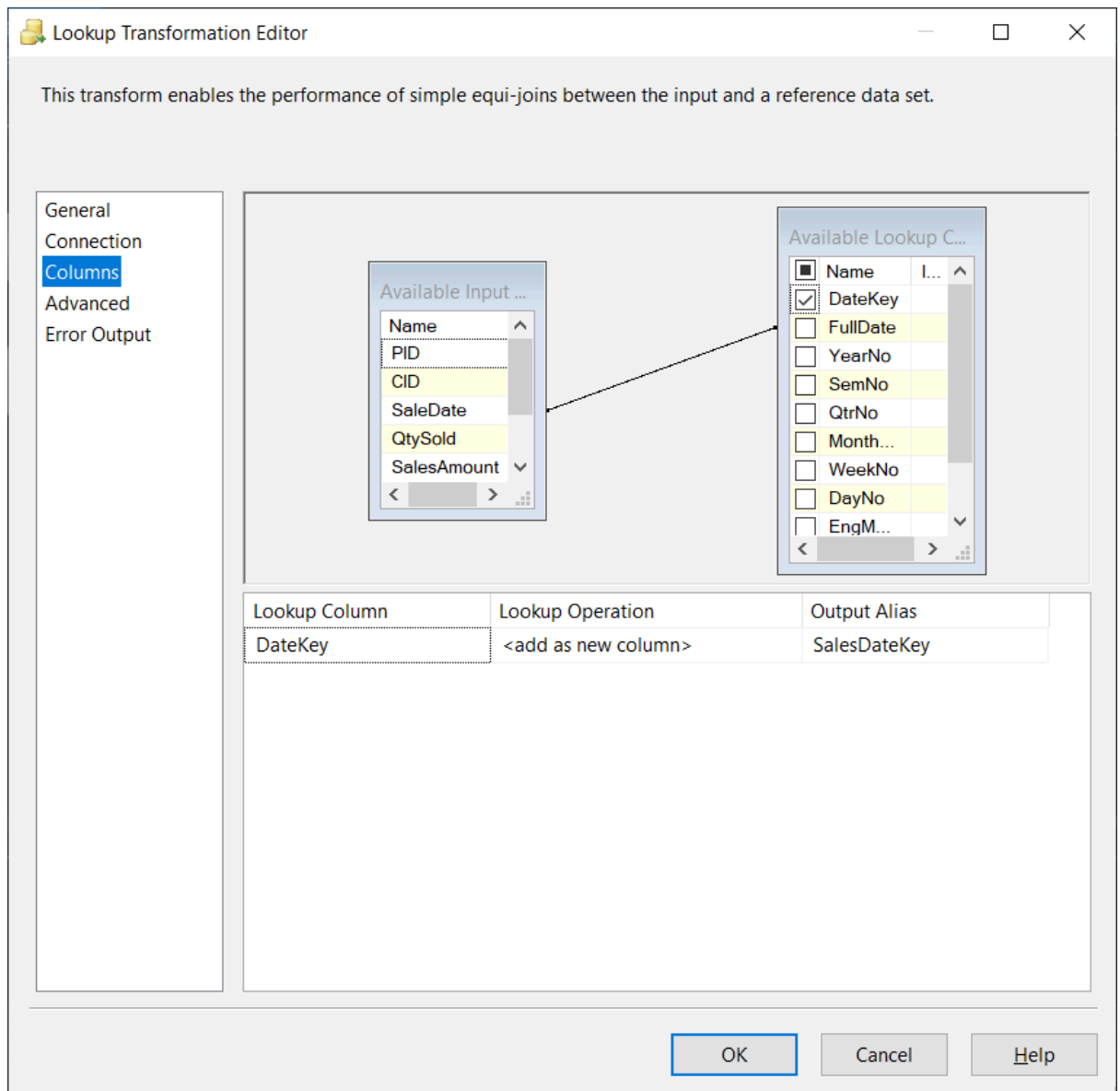
### 5.3.3. TẢI DỮ LIỆU BÁN HÀNG VÀO DATAWARE HOUSE

Task “Data Flow” sử dụng Toolbox “Source Assistant” để chỉ định bảng BeverageStage.StageSales làm bảng nguồn và “Destination Assistant” chỉ định bảng BeverageDW.FactSales làm bảng đích tương tự như các quá trình tải dữ liệu khách hàng và dữ liệu sản phẩm.

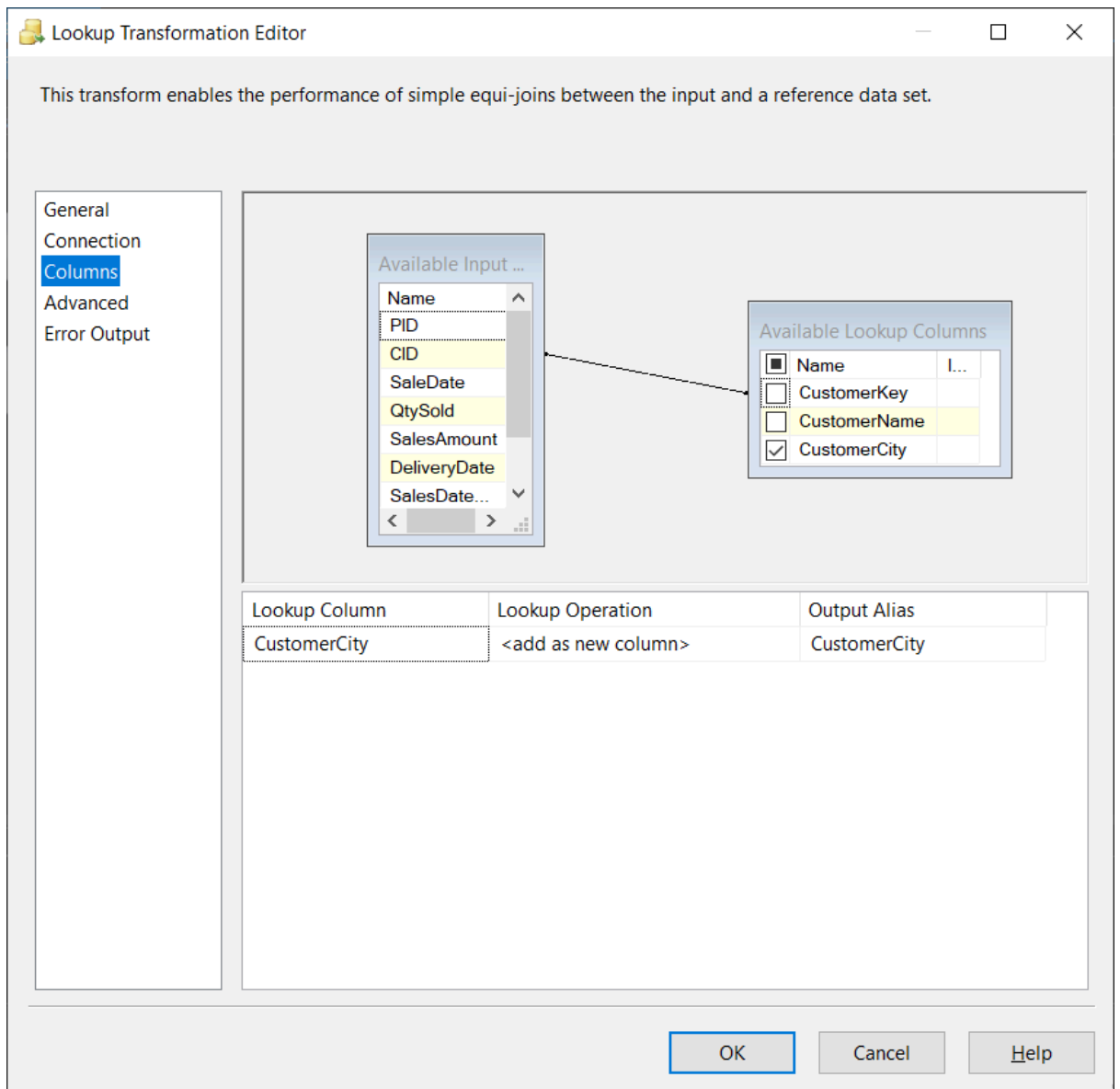
Sử dụng các Toolbox “Lookup” để lấy giá trị “DateKey” từ bảng DimTime, giá trị “CustomerCity” từ DimCustomer và giá trị “LocationKey” từ bảng DimLocation làm các cột mới cho bảng FactSales.



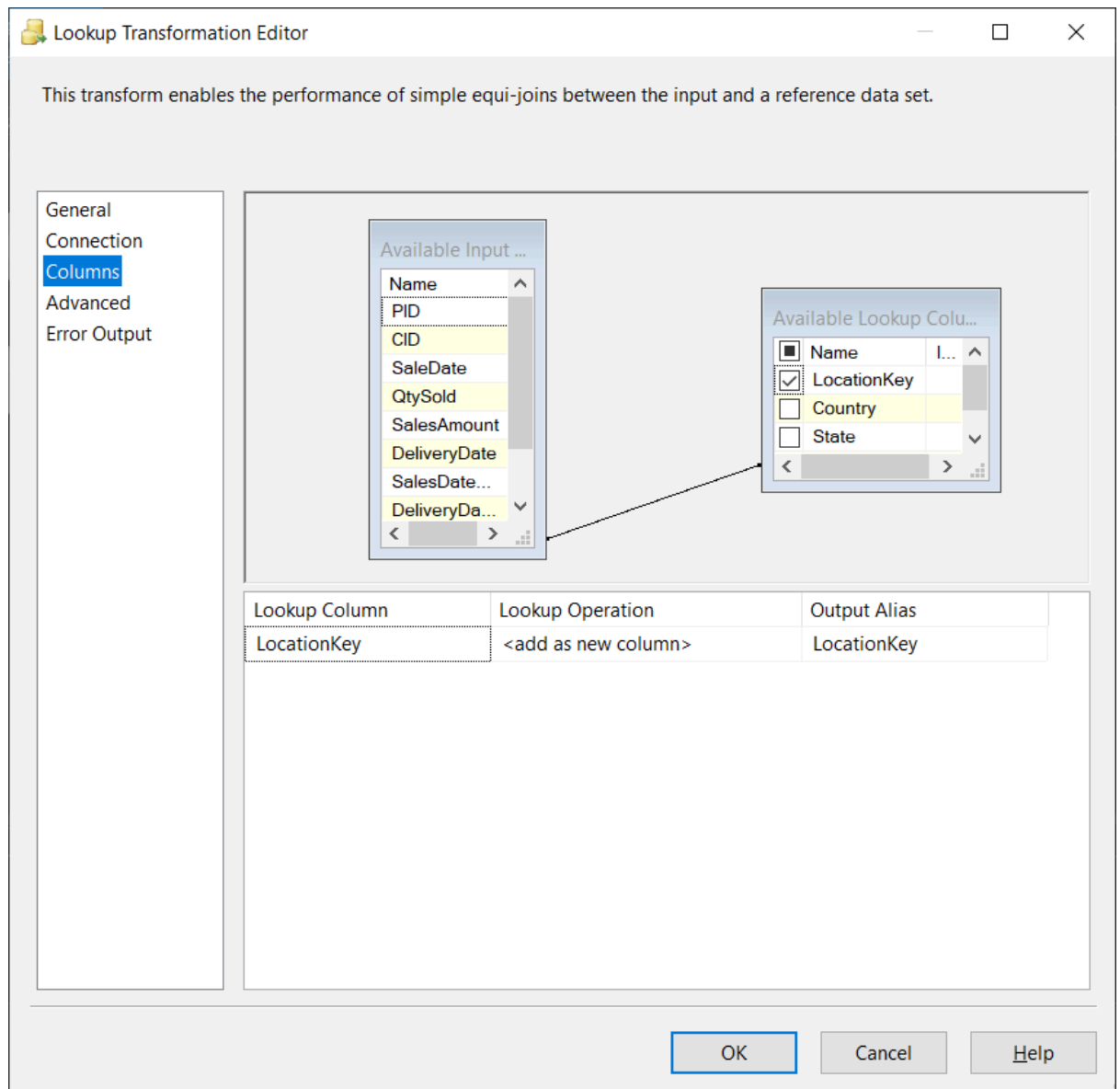
Hình 2.5.3.3.1. Cấu hình task “Data Flow”



Hình 2.5.3.3.2. Cấu hình ToolBox “Lookup” lấy “DateKey” cho “SalesDate” và “DeliveryDate”



Hình 2.5.3.3.3. Cấu hình Toolbox “Lookup” lấy dữ liệu “CustomerCity”



Hình 2.5.3.3.4. Cấu hình Toolbox “Lookup” lấy giá trị “LocationKey”

Sau khi chạy thành công, bảng FactSales có dữ liệu như hình bên dưới.



```

1  /***** Script for SelectTopNRows command from SSMS *****/
2  SELECT TOP (1000) [ProductKey]
3      , [CustomerKey]
4      , [SaleDateKey]
5      , [QtySold]
6      , [SalesAmount]
7      , [DeliveryDateKey]
8      , [LocationKey]
9  FROM [BeverageDW].[dbo].[FactSales]

```

107 %

Results Messages

	ProductKey	CustomerKey	SaleDateKey	QtySold	SalesAmount	DeliveryDateKey	LocationKey
1	10	10	20190103	30	6000.00	20190106	112
2	15	12	20190104	300	24000.00	20190107	113
3	17	22	20190115	30	1800.00	20190118	109
4	21	26	20190116	80	2400.00	20190119	108
5	28	30	20190117	300	6000.00	20190120	112
6	29	36	20190118	150	450000.00	20190121	190
7	20	33	20190123	30	750000.00	20190126	185
8	11	10	20190124	30	1800.00	20190127	112
9	1	60	20190125	30	18000.00	20190128	167
10	2	10	20190126	30	300.00	20190129	112
11	4	10	20190203	30	810000.00	20190206	112
12	5	12	20190204	100	10000.00	20190207	113
13	12	22	20190215	30	1050.00	20190218	109
14	8	26	20190216	80	8000.00	20190219	108
15	9	30	20190217	300	30000.00	20190220	112
16	3	36	20190218	150	1800000.00	20190221	190
17	1	33	20190223	30	750000.00	20190226	185
18	6	10	20190224	50	5000.00	20190227	112
19	3	60	20190225	100	20000.00	20190228	167
20	7	10	20190226	90	900.00	20190301	112

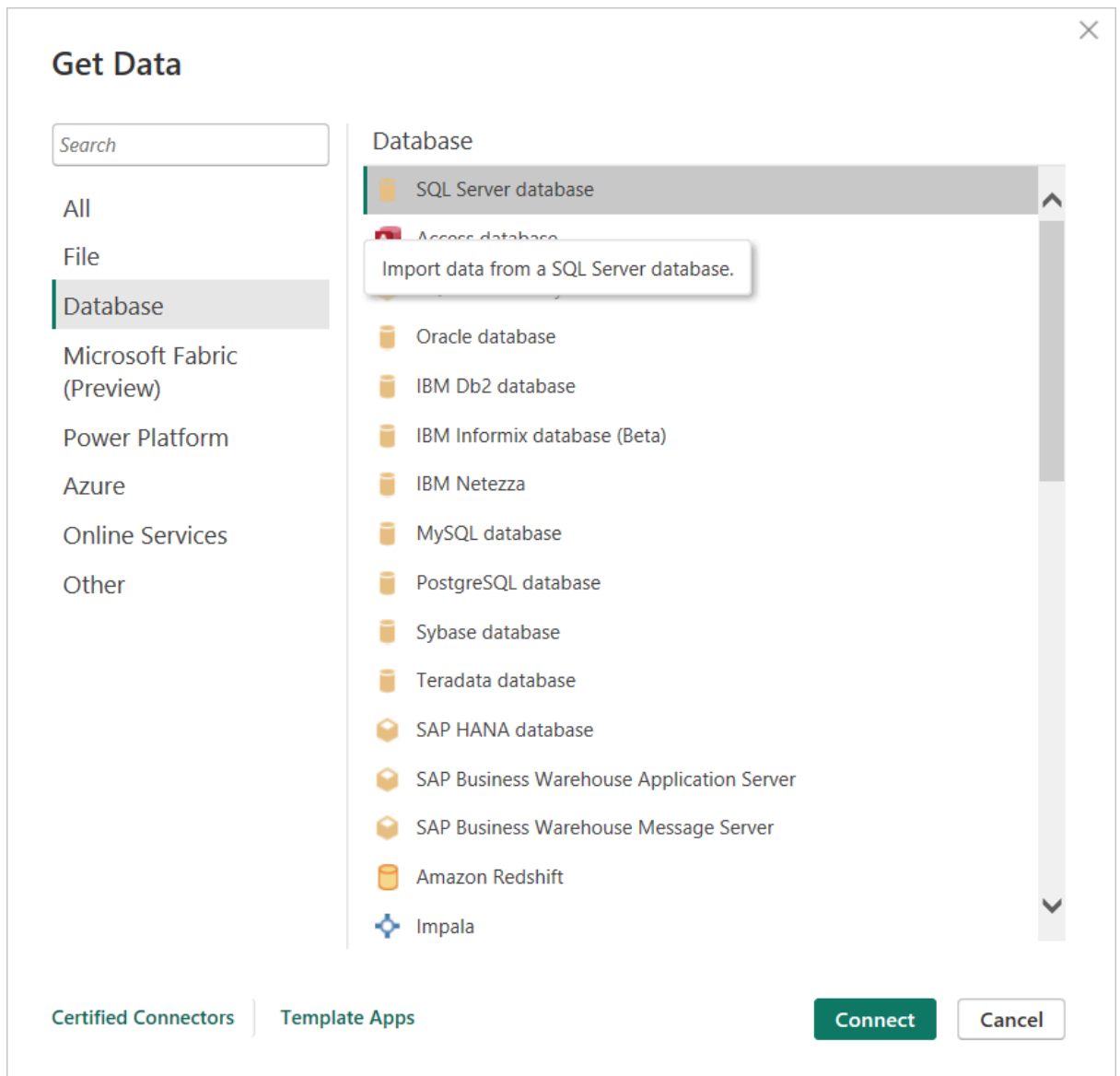
Hình 2.5.3.3.5. Một số dòng dữ liệu của bảng FactSales

## 5.4. TRỰC QUAN HÓA DỮ LIỆU BẰNG POWER BI

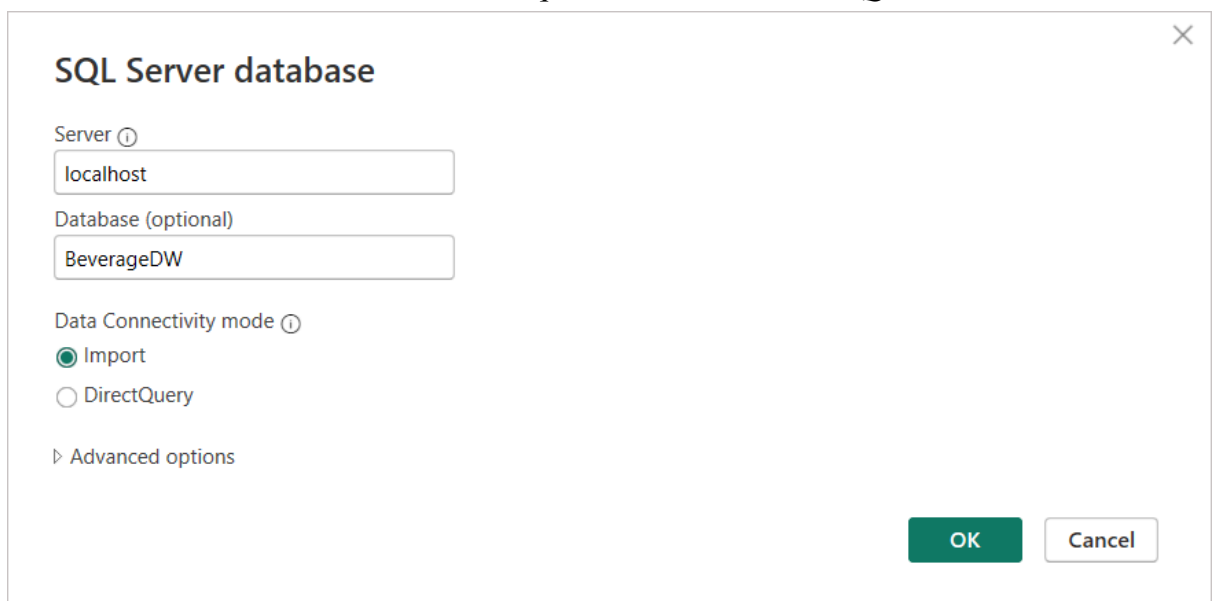
Sau khi thực hiện ETL dữ liệu vào Data Warehouse, đề tài sử dụng dữ liệu của cơ sở dữ liệu BeverageDW để trực quan hóa thành các biểu đồ và số liệu thống kê.

### 5.4.1. LẤY DỮ LIỆU TỪ DATA WAREHOUSE

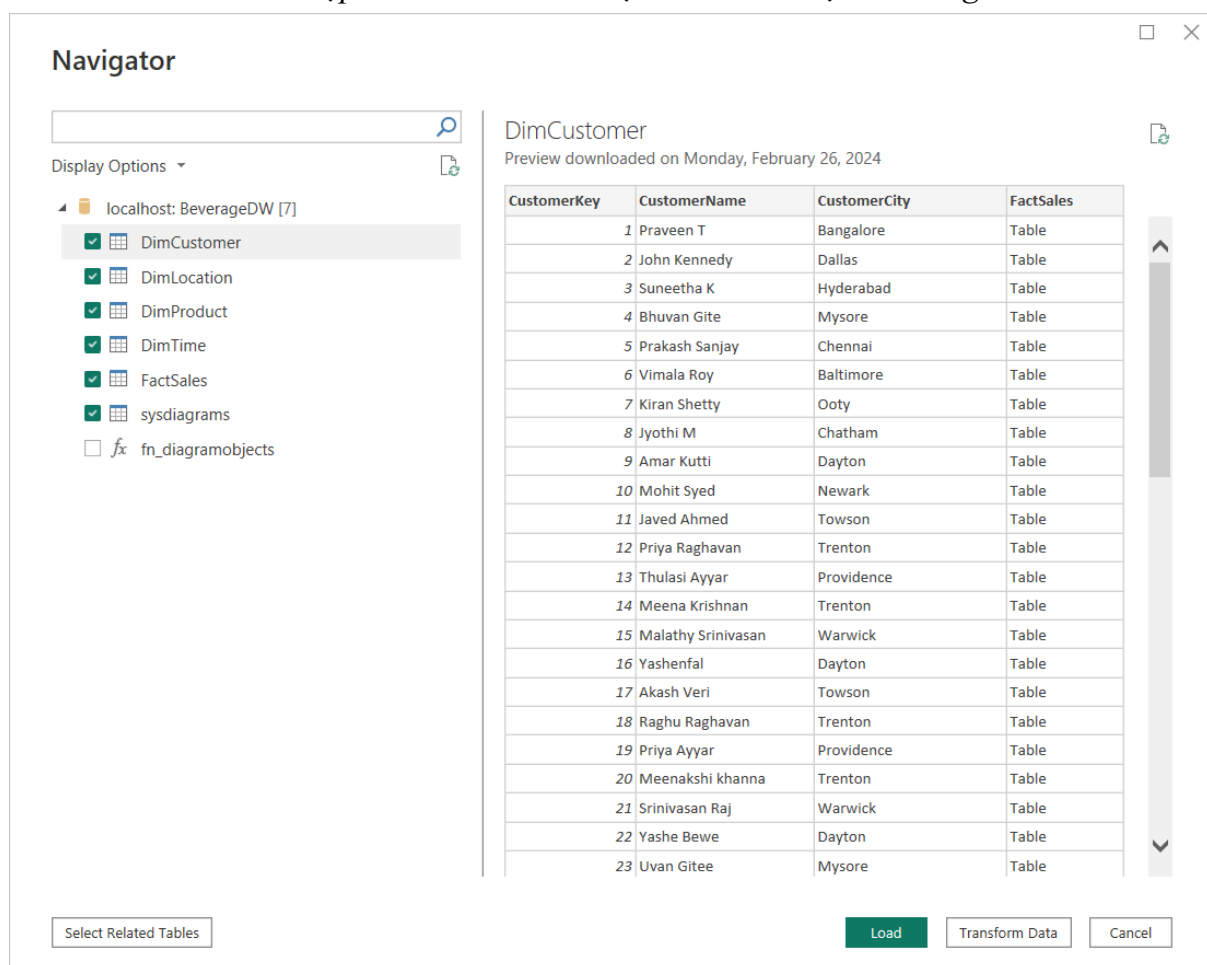
Chọn nguồn dữ liệu là cơ sở dữ liệu BeverageDW được lưu trữ trong SQL Server bằng chức năng “Get Data”.



*Hình 2.5.4.1.1. Kết nối đến hệ quản trị cơ sở dữ liệu SQL Server*



Hình 2.5.4.1.2. Nhập tên server và chỉ định cơ sở dữ liệu BeverageDW

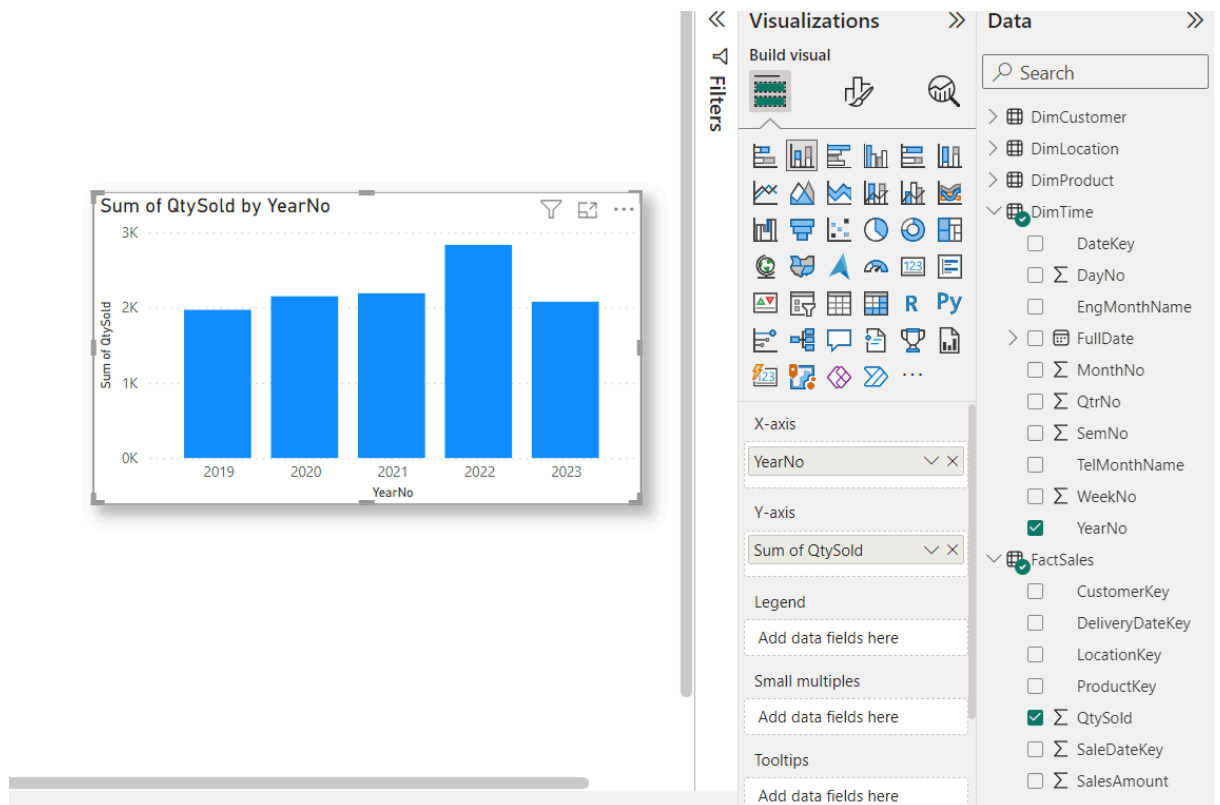


Hình 2.5.4.1.3. Chọn các bảng dữ liệu phục vụ cho quá trình thống kê

## 5.4.2. TẠO DASHBOARD

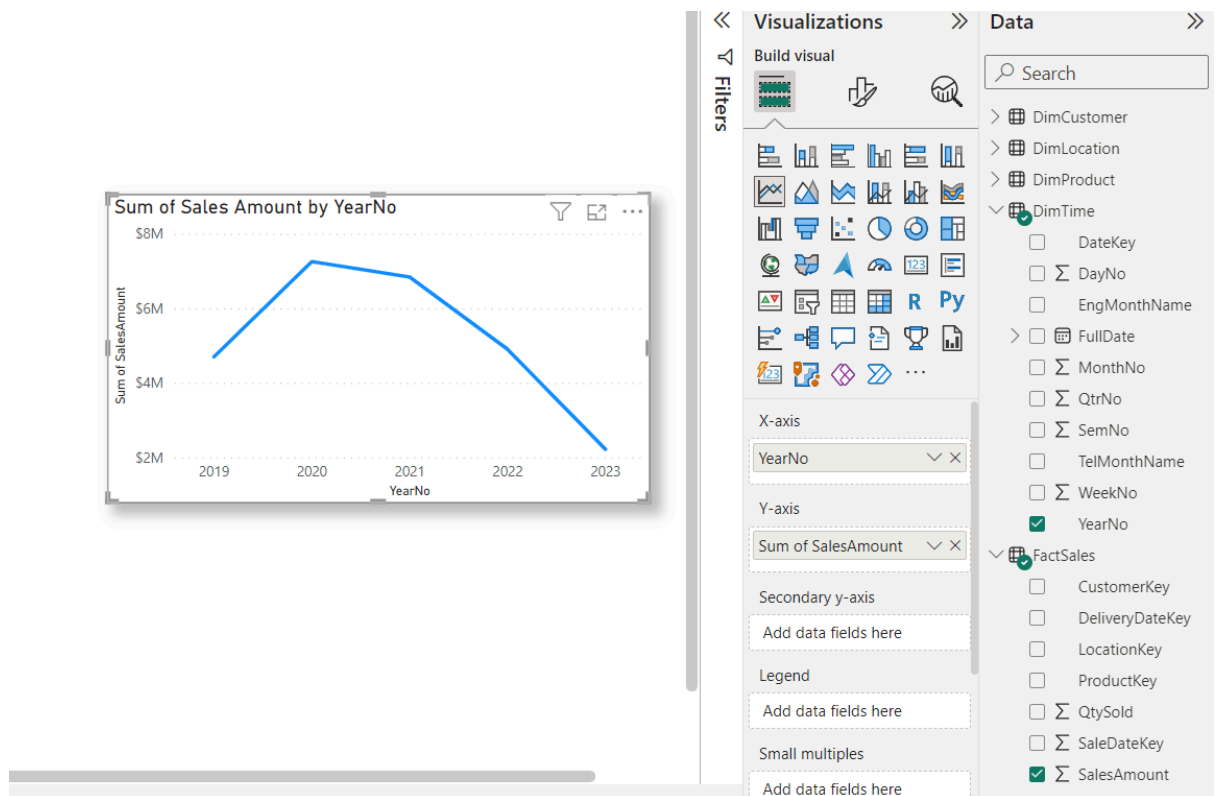
Dashboard là một bảng kê bao gồm các biểu đồ được trực quan hóa dựa vào các số liệu được lưu trữ trong Data Warehouse giúp người dùng có được góc nhìn tổng quan về hoạt động của doanh nghiệp một cách nhanh chóng.

Sử dụng chức năng “Stacked column chart” tạo biểu đồ cột để biểu diễn số lượng sản phẩm bán được qua từng năm từ 2019 đến 2023. Trục x là giá trị “YearNo” được lấy từ bảng DimTime, trục y được tính bằng tổng giá trị “QtySold” trong bảng FactSales.



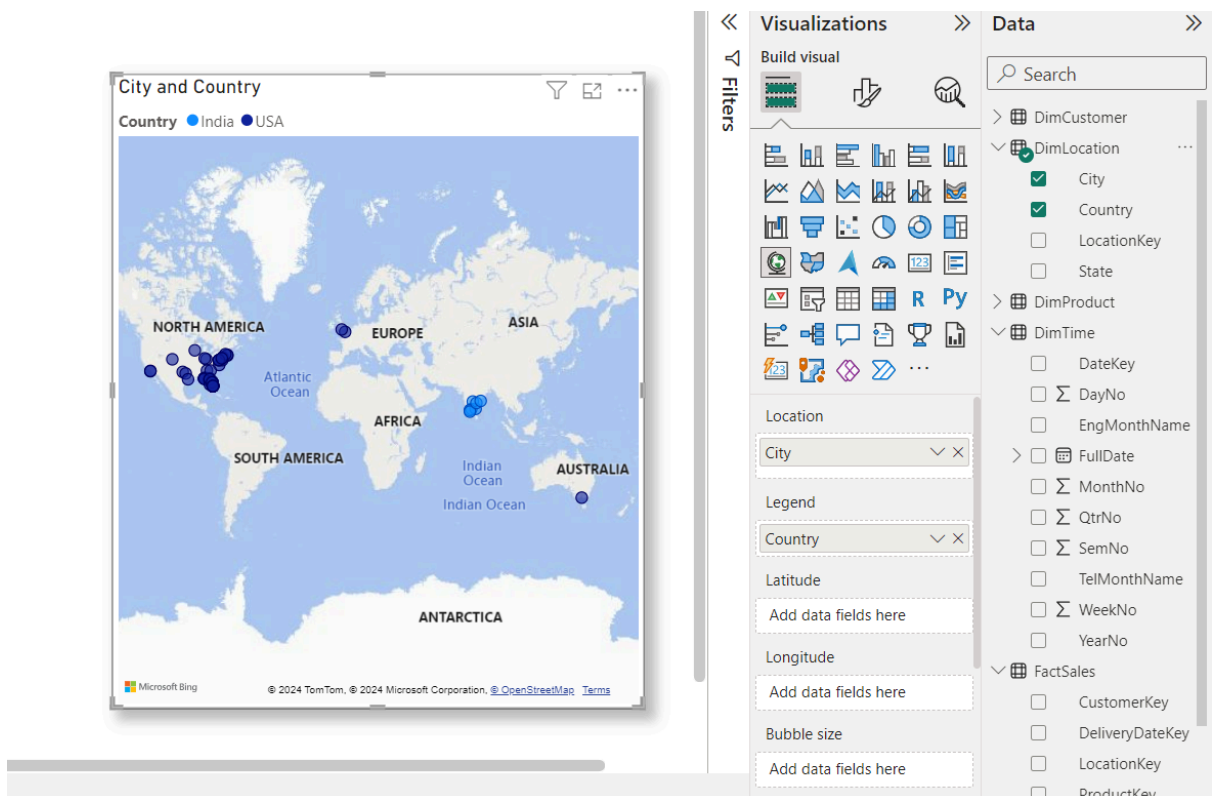
Hình 2.5.4.2.1. Thiết kế và dữ liệu của biểu đồ “Sum of QtySold by YearNo”

Biểu đồ “Sum of Sales Amount by YearNo” là biểu đồ đường (Line chart) dùng để biểu diễn doanh số của doanh nghiệp qua từng năm hoạt động. Trục x là giá trị “YearNo” được lấy từ bảng DimTime, trục y được tính bằng tổng giá trị “SalesAmount” trong bảng FactSales.



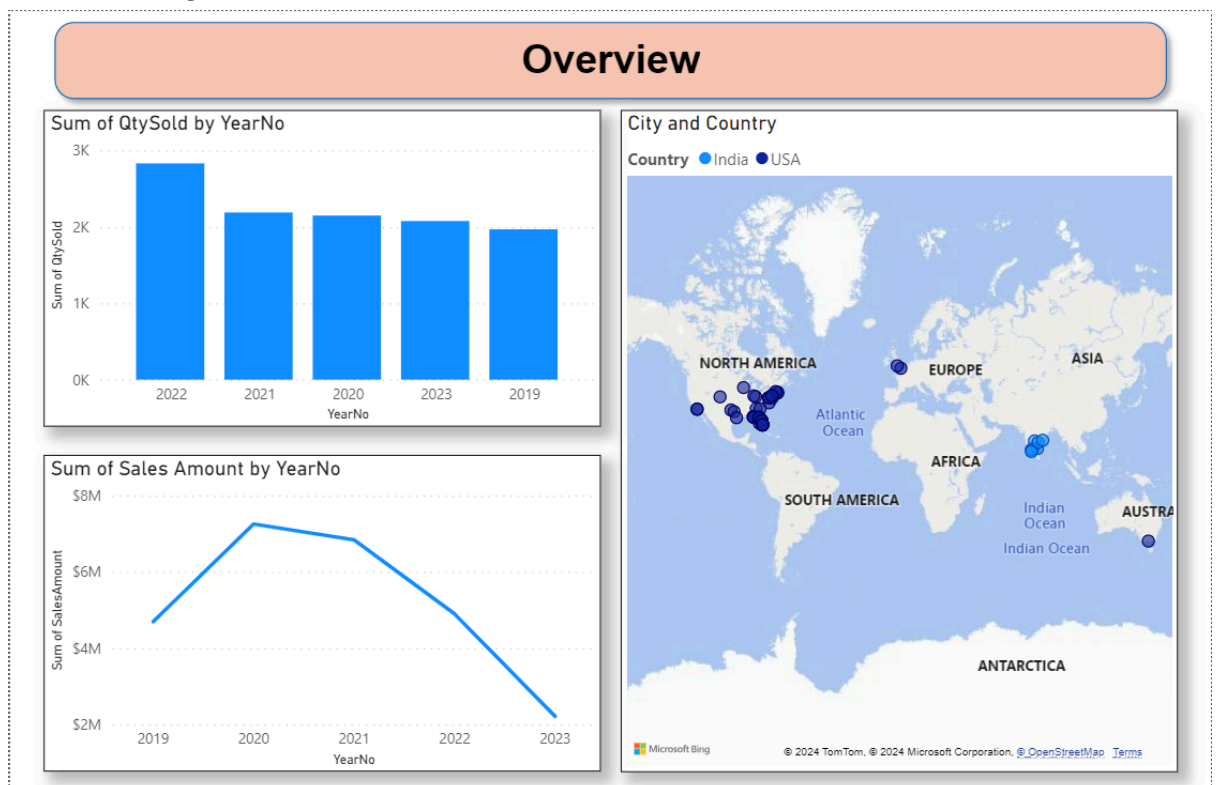
Hình 2.5.4.2.2. Thiết kế và dữ liệu của biểu đồ “Sum of Sales Amount by YearNo”

Biểu đồ “City and Country” là một “map” biểu diễn địa chỉ của khách hàng bao gồm quốc gia và thành phố để xác định thị trường của doanh nghiệp. Dữ liệu được lấy từ cột “City” và “Country” của bảng DimLocation.



Hình 2.5.4.2.3. Thiết kế và dữ liệu của biểu đồ “City and Country”

Sau khi tổng hợp các biểu đồ trên, Dashboard của doanh nghiệp có dạng như hình bên dưới.

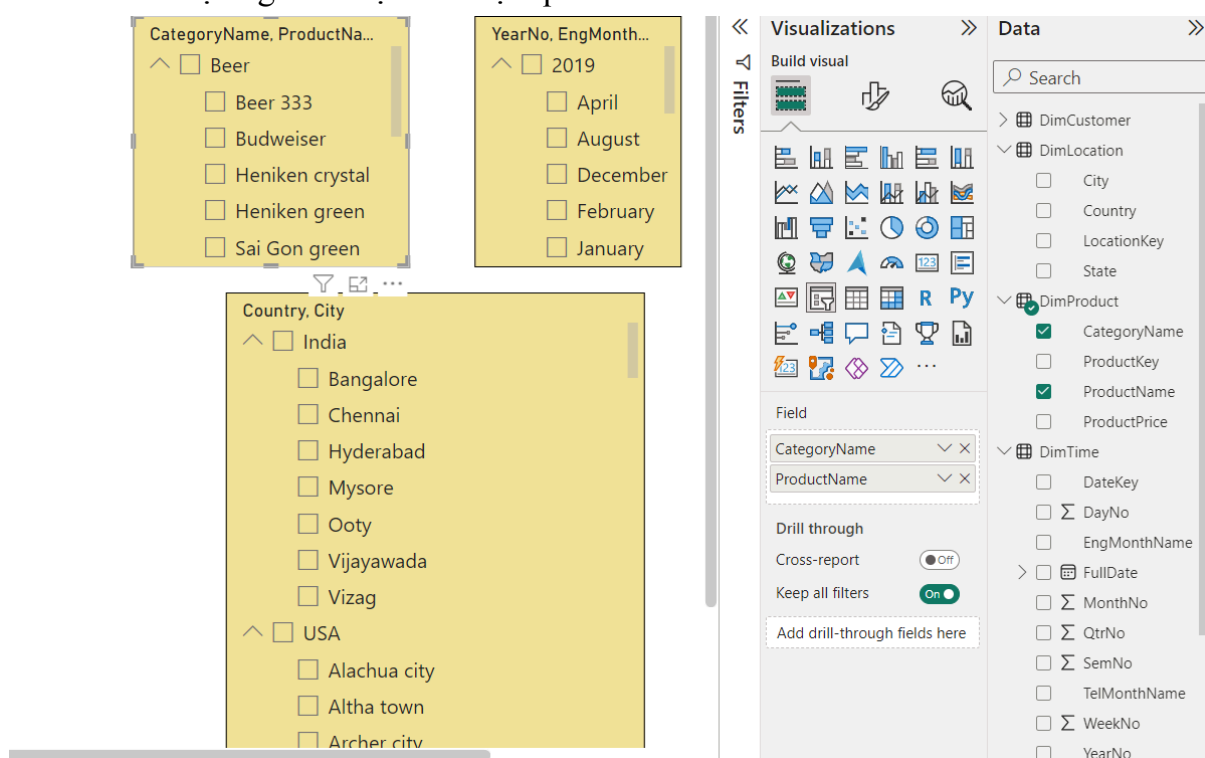


Hình 2.5.4.2.4. Dashboard biểu diễn tổng quan số liệu của doanh nghiệp

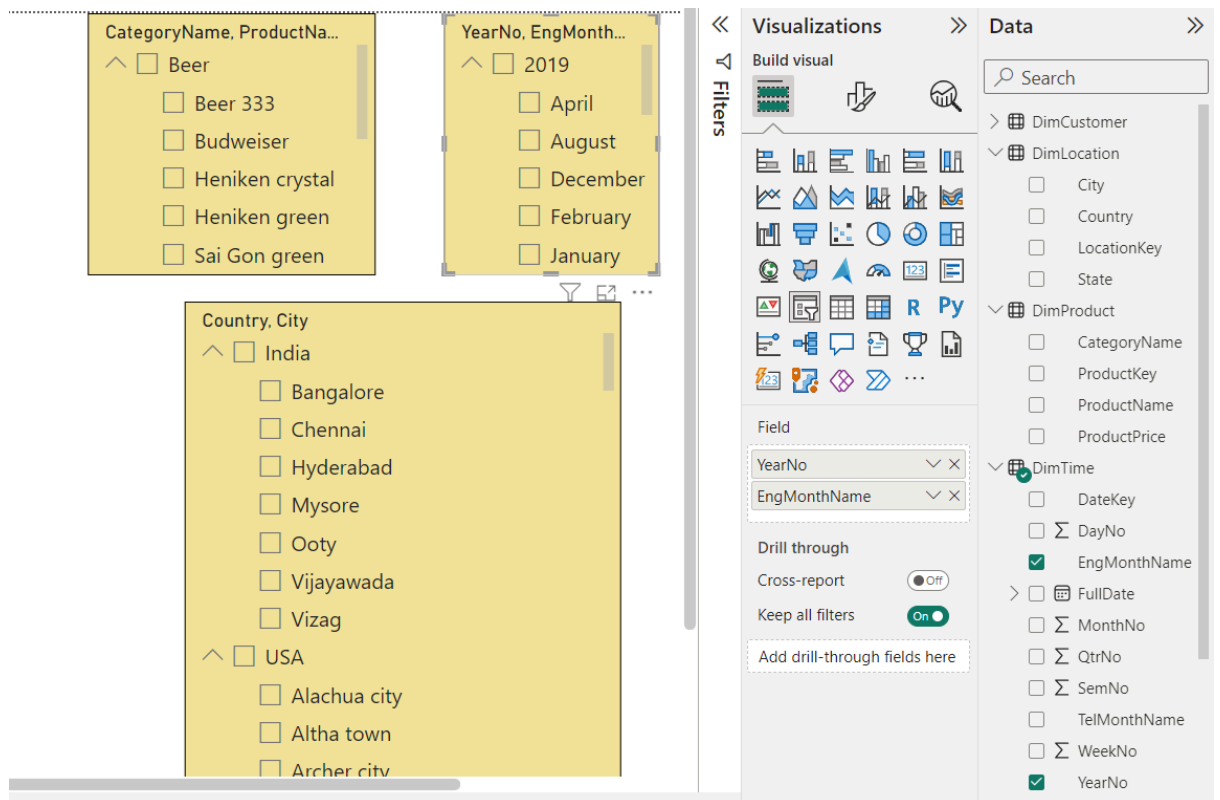
### 5.4.3. TẠO SALES REPORT

Sales report là báo cáo quản lý bán hàng dùng để thống kê các số liệu thu thập được từ hoạt động của bộ phận bán hàng trong doanh nghiệp bao gồm tổng doanh số và tổng sản phẩm bán được. Các biểu đồ và số liệu này có thể thay đổi dựa vào bộ lọc như lọc theo sản phẩm, lọc theo tháng, năm,...

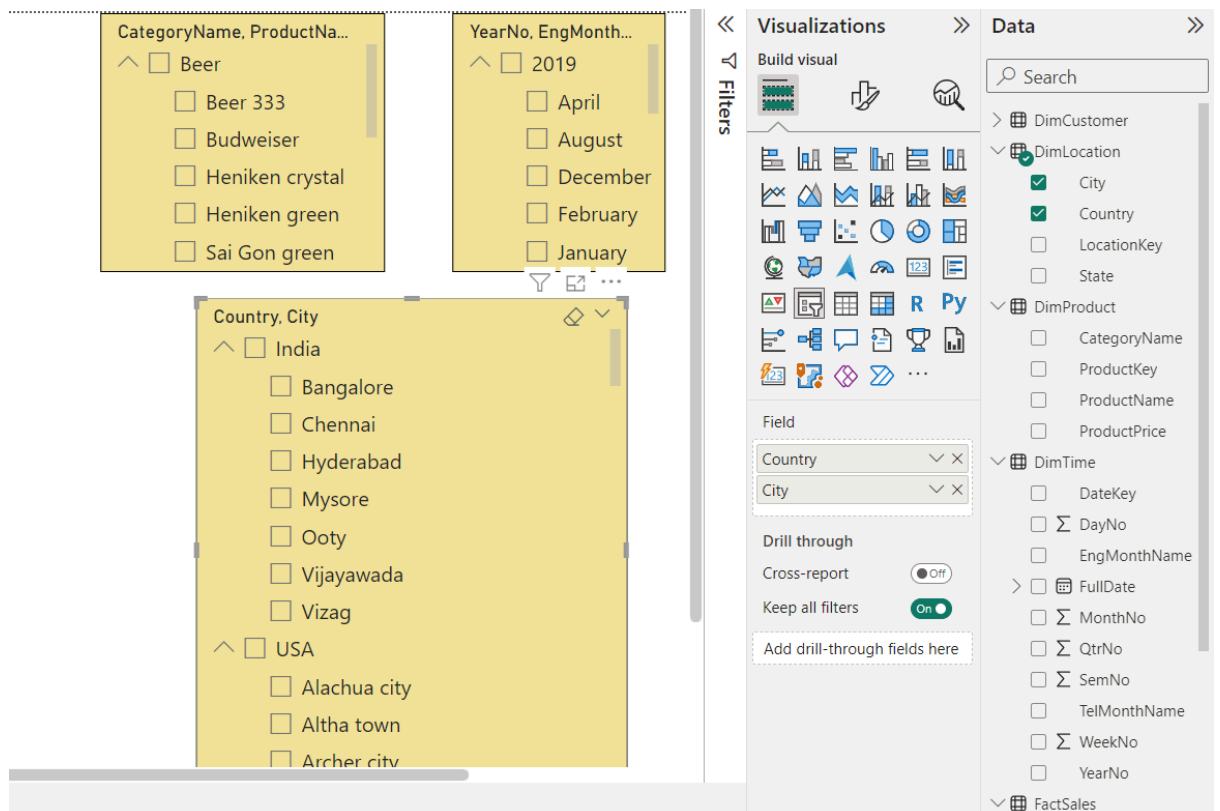
Sử dụng “Slicer” để tạo bộ lọc theo các thuộc tính CategoryName - ProductName, YearNo - EngMonth, Country - City. Cụ thể thiết kế và dữ liệu nguồn được thể hiện qua các hình bên dưới.



Hình 2.5.4.3.1. Thiết kế và dữ liệu của slicer “CategoryName, ProductName”



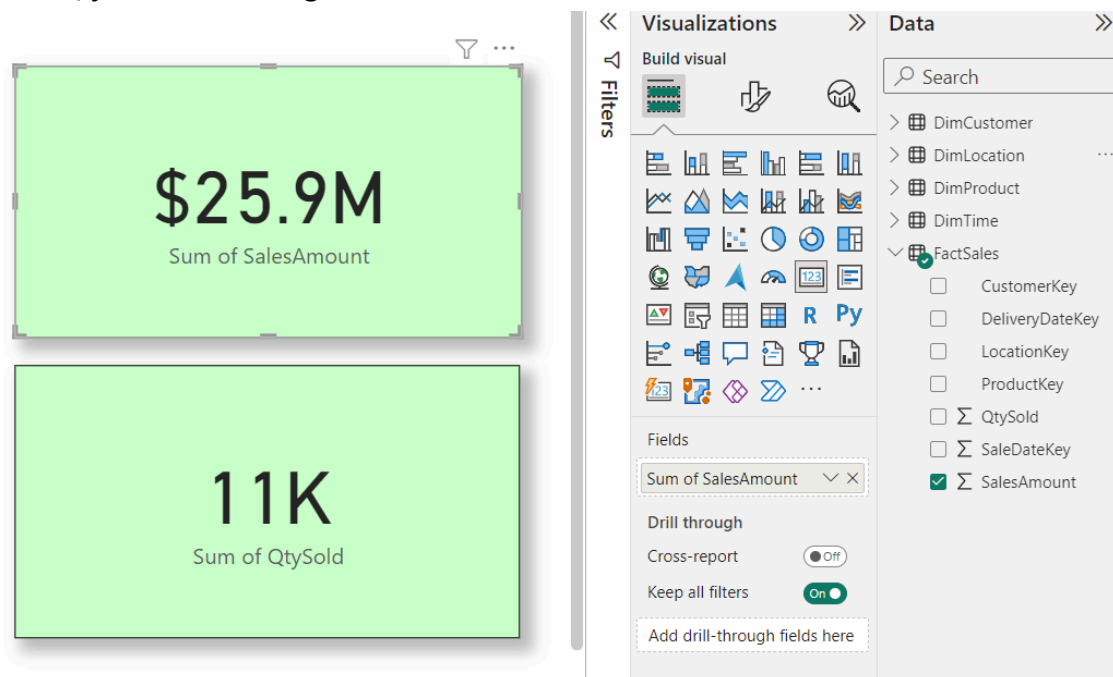
Hình 2.5.4.3.2. Thiết kế và dữ liệu của slicer “YearNo, EngMonth”



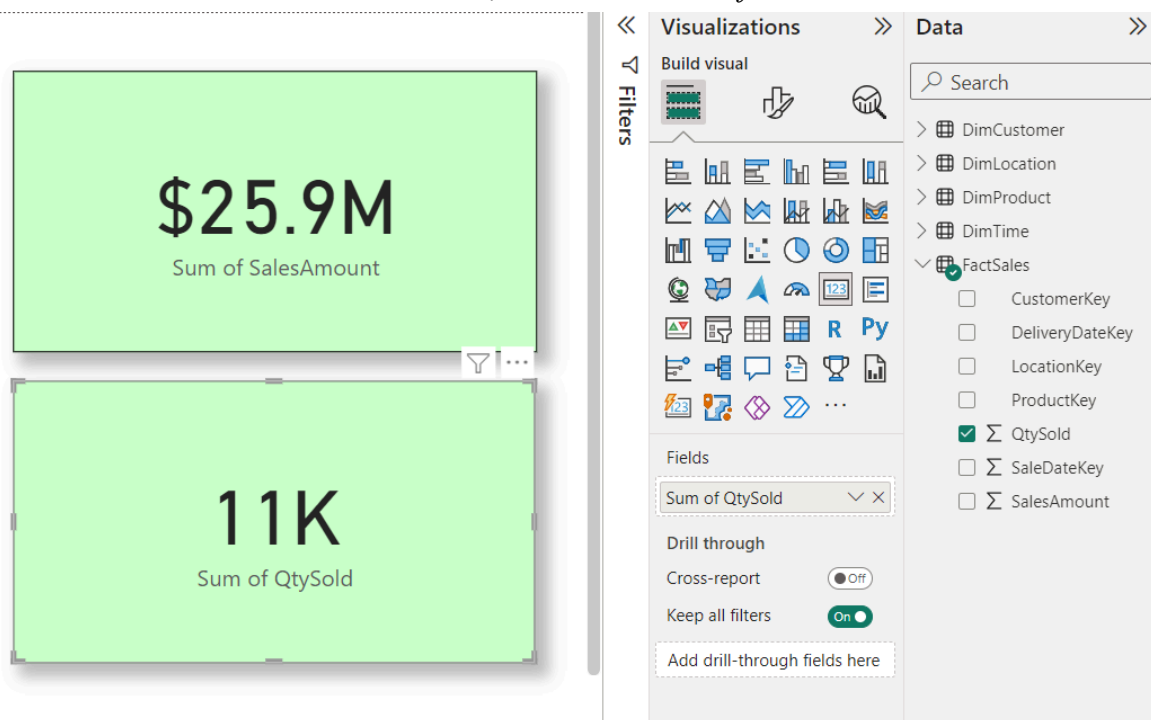
Hình 2.5.4.3.3. Thiết kế và dữ liệu của slicer “Country, City”



Sử dụng chức năng “Card” để tạo thẻ biểu diễn tổng doanh số và tổng số lượng sản phẩm bán ra. Dữ liệu được lấy từ cột “SalesAmount” và “QtySold” của bảng FactSales.

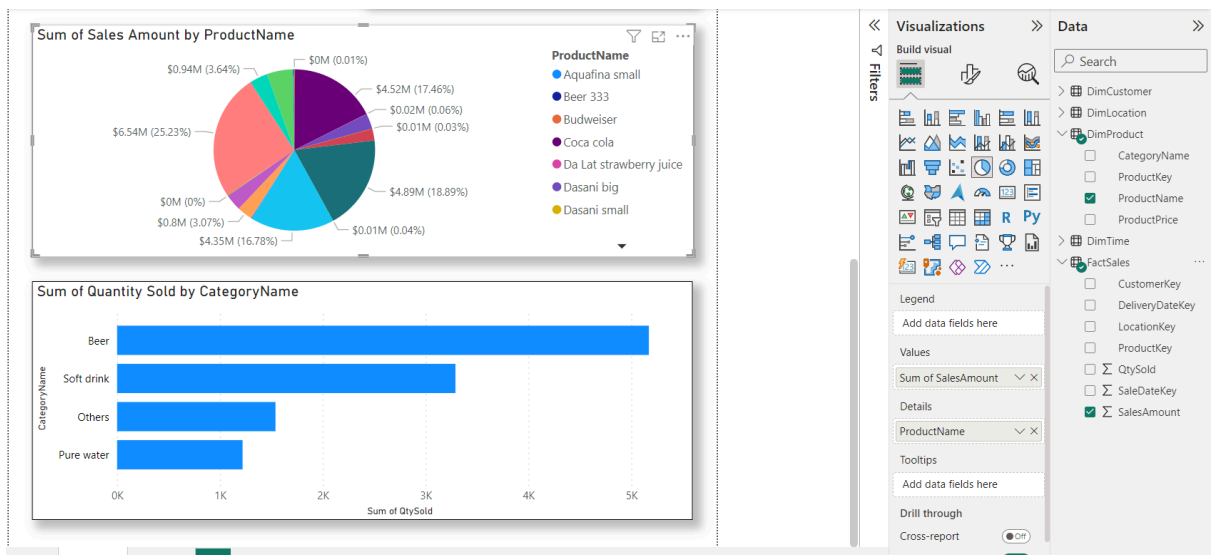


Hình 2.5.4.3.4. Thiết kế và dữ liệu của thẻ “Sum of SalesAmount”

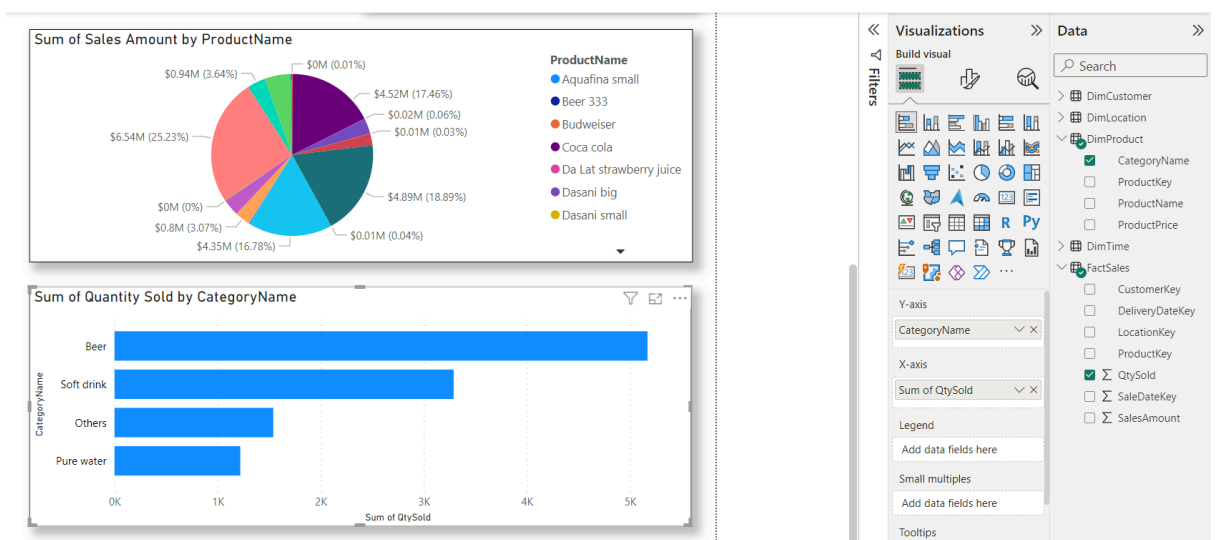


Hình 2.5.4.3.5. Thiết kế và dữ liệu của thẻ “Sum of QtySold”

Sử dụng “Pie Chart” để tạo biểu đồ tròn biểu diễn doanh số theo từng sản phẩm và một biểu đồ cột biểu diễn số lượng sản phẩm bán ra theo danh mục sản phẩm.

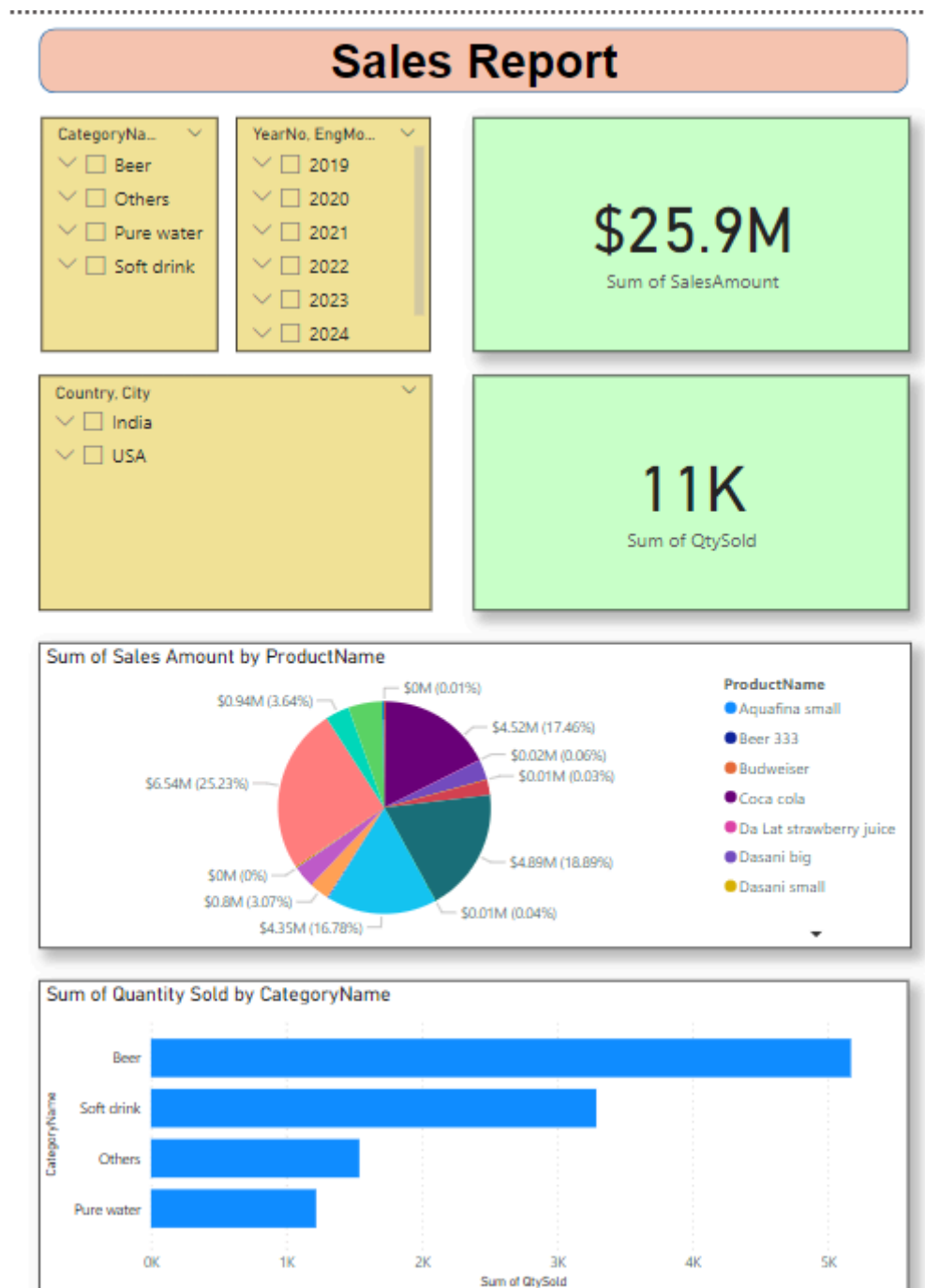


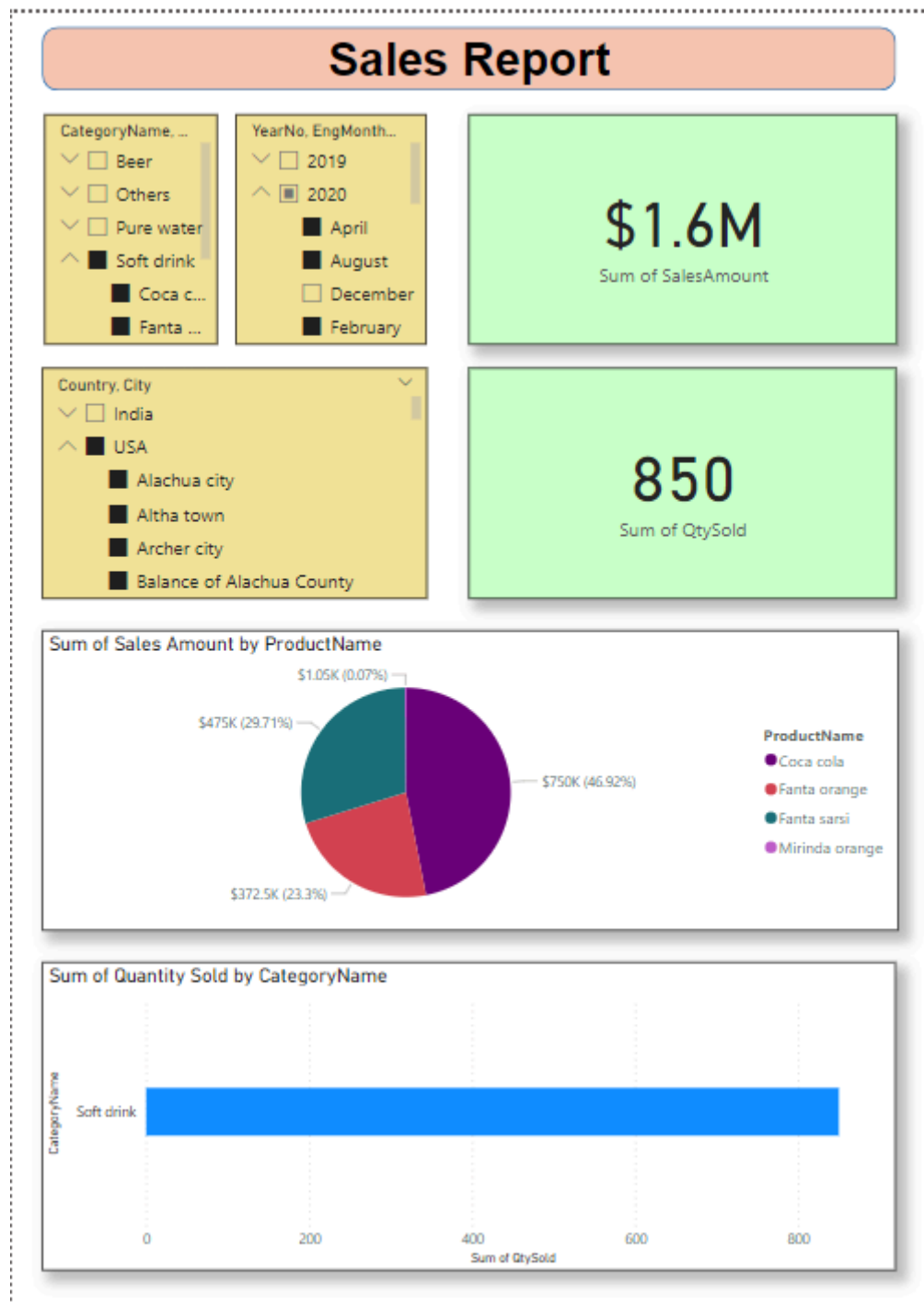
Hình 2.5.4.3.6. Thiết kế và dữ liệu biểu đồ “Sum of Sales Amount by ProductName”



Hình 2.5.4.3.7. Thiết kế và dữ liệu biểu đồ “Sum of Quantity Sold by Category”

Sau khi tổng hợp các biểu đồ trên, ta có báo cáo quản lý bán hàng hoàn chỉnh như hình bên dưới.





Hình 2.5.4.3.9. Báo cáo sau khi đã lọc một số mục

## KẾT LUẬN

Trong bối cảnh các doanh nghiệp phân phối nước giải khát cần lưu trữ rất nhiều thông tin dữ liệu về hoạt động kinh doanh của mình nhưng chưa có một giải pháp cụ thể, đề tài quyết định nghiên cứu xây dựng Data Warehouse với mục đích giúp họ quản lý dữ liệu kinh doanh hiệu quả hơn và dễ dàng sử dụng dữ liệu để phân tích, đánh giá và đưa ra các thống kê báo cáo cũng như những dự đoán sắp tới cho phương hướng phát triển của doanh nghiệp.

Sau khi nghiên cứu, nhóm nhận thấy SSIS là một công cụ xây dựng kho dữ liệu “no code”. Đây là một phương án sử dụng công cụ để tạo kho dữ liệu dành cho đối tượng không có kinh nghiệm lập trình hoặc không muốn sử dụng code trong quá trình tạo ra kho dữ liệu.

Đề tài vẫn còn hạn chế ở giai đoạn sao chép dữ liệu vào vùng Staging. Mỗi lần cập nhật dữ liệu phải chạy câu lệnh “Truncate” để xóa sạch dữ liệu cũ rồi mới thêm lại toàn bộ dữ liệu mới vào. Sau nhiều lần cập nhật, các dữ liệu cũ sẽ nhiều lên cộng với lượng dữ liệu lớn thêm vào sẽ làm chậm quá trình cập nhật. Ngoài ra, các quy trình thực hiện đa số vẫn còn thủ công chưa tự động hóa hoàn toàn.

Để phát triển đề tài này, ta có thể sử dụng “Schedules” của SSIS để cập nhật dữ liệu định kỳ kèm theo đó là kiểm soát các lần ETL sau để phát hiện dữ liệu cũ đã được lưu, từ đó tiến hành xử lý để bỏ qua dữ liệu đó và chỉ thêm những dữ liệu mới chưa từng được lưu.

## ❖ TÀI LIỆU THAM KHẢO

- [1] “Phong Nguyễn”, “Kho dữ liệu - Chiếc dạ dày của hệ thống BI”, 8/3/2021. [Trực tuyến]. Địa chỉ: <https://ierp.vn/data-warehouse-la-gi/>. [Truy cập 10/3/2024].
- [2] “Hieu Nguyen”, “Data Warehouse là gì? Đặc điểm, thành phần, cách thức hoạt động và ứng dụng”, 24/5/2023. [Trực tuyến]. Địa chỉ: <https://www.toponseek.com/blogs/data-warehouse-la-gi/>. [Truy cập 10/3/2024].
- [3] “The Investopedia team”, “What Is a Data Warehouse? Warehousing Data, Data Mining Explained”, 14/9/2022. [Trực tuyến]. Địa chỉ: <https://www.investopedia.com/terms/d/data-warehousing.asp>. [Truy cập: 10/3/2024].
- [4] “Dương Thị Tâm”, “Tổng Quan Và Cách Cài Đặt Của MS SQL Server”, 24/6/2018. [Trực tuyến]. Địa chỉ: <https://viblo.asia/p/tong-quan-va-cach-cai-dat-cua-ms-sql-server-3P0lPzJGKox>. [Truy cập 10/3/2024].
- [5] “ITNavi”, “SQL Server là gì? Mục đích của việc sử dụng SQL Server”, 20/10/2020. [Trực tuyến]. Địa chỉ: <https://itnavi.com.vn/blog/sql-server>. [Truy cập 10/3/2024].
- [6] “LANIT JSC”, “SSIS là gì? Thông tin chi tiết về Công Dụng, Thành Phần Chính”, 19/10/2023. [Trực tuyến]. Địa chỉ: <https://lanit.com.vn/ssis-la-gi.html>. [Truy cập 10/3/2024].
- [7] “TACA”, “CÔNG CỤ POWER BI LÀ GÌ? ỨNG DỤNG POWER BI TRONG VIỆC LẬP BÁO CÁO QUẢN TRỊ DOANH NGHIỆP”, 27/9/2022. [Trực tuyến]. Địa chỉ: <https://taca.com.vn/power-bi/>. [Truy cập 10/3/2024].