**BÁO CÁO**

DEP302 – Assignment 01

Họ và tên: Nguyễn Hải Dương

Mã Sinh viên: FX16133

1. **Thiết kế ERD**

Data warehouse được thiết kế theo kiến trúc Centralized Data Warehouse. Bao gồm các bảng như ERD dưới: A screenshot of a computer

Description automatically generated

Các câu lệnh SQL để xây dựng cơ sở dữ liệu

File: CreateTableQuery.sql

1. Tạo các bảng cho các dimensions.

A screenshot of a computer

Description automatically generatedA screenshot of a computer code

Description automatically generated

1. Tạo bảng cho fact table cùng các khóa phụ A screen shot of a computer program

   Description automatically generatedA screenshot of a computer program

   Description automatically generated
2. **Xác định các truy vấn nghiệp vụ:**

Các truy vấn nghiệp vụ được sử dụng cho cơ sở dữ liệu của bài

1. Số lượng thú nuôi chưa được tiêm vaccine và chưa được sổ giun, phân theo từng khu vực (state)
2. Số lượng thú cưng thuần chủng (Không lai giữa 2 giống khác nhau) phân thao từng loài
3. Phí nhận nuôi trung bình theo kích thước trưởng thành
4. Phí nhận nuôi trung bình theo giống (Bao gồm cả giống 1 và 2 nếu có)
5. **Xây dựng ETL cho Dimension tables:**
6. Quy trình xây dựng ETL cho các Dimension table:

Bảng PetType\_Dimension

A diagram of a process

Description automatically generated

Gồm 4 bước:

1. Nhập dữ liệu từ file CSV

A close up of a logo

Description automatically generated

1. A screenshot of a computer

   Description automatically generatedSort dữ liệu theo type nhằm loại bỏ các giá trị bị trùng
2. Lookup – Tìm/So sánh các dữ liệu đã có sẵn trong bảng PetType\_Dim

A screenshot of a computer

Description automatically generatedChọn “Redirected rows to match output” để sử dụng những dữ liệu chưa có trong bảng PetType\_Dim và nhập vào bảng này ở bước sau

A screenshot of a computer

Description automatically generatedTạo Connection với bảng PetType\_dim trong Database DEP302\_ASM1

A screenshot of a computer

Description automatically generated

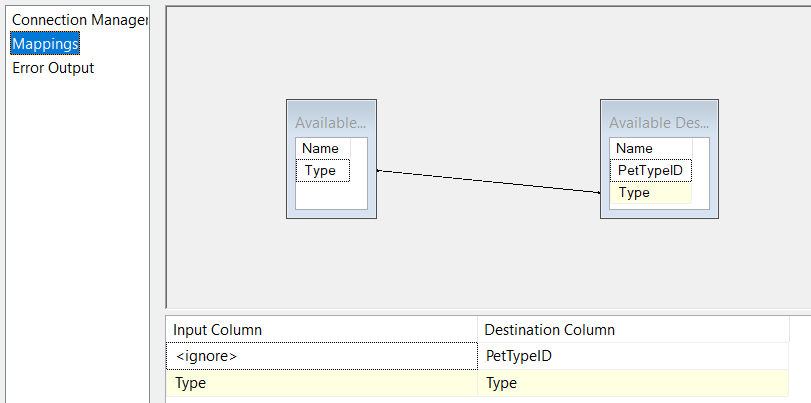
So sánh Type trong dữ liệu sau khi sort với dữ liệu look up trong bảng PetType\_Dim

Kết nối dữ liệu của lookup theo dạng “Lookup no match output” với Destination

1. Truyền dữ liệu vào bảng PetType\_Dim trong cơ sở dữ liệu:

A screenshot of a computer

Description automatically generatedSử dụng OLE DB Destination để kết nối với bảng PetType\_Dim



Mapping Type dữ liệu Type với Type trong bảng PetType\_Dim

Tương tự với các dimension tables:

* FurLength\_Dim
* Gender\_Dim
* MaturittySize\_Dim
* Health\_Dim
* State\_Dim
* Breed\_Dim
* Color\_Dim

1. Thiết lập các Data Flow chạy đồng thời sử dụng Sequence Container:

Sau khi hoàn thành thiết lập cho các Dimension tables, ta được các các Data Flow như hình dưới.  
Ta sử dụng Sequence Container để nhóm các Data Flow cho việc nhập dữ liệu các Dimension tables nhằm để cho các Data flow này được xử lý song song.

A screenshot of a computer

Description automatically generated

1. **Xây dựng ETL cho Fact table**
2. Tạo 1 Data Flow mới cho việc nhập dữ liệu Fact table:

A close-up of a computer screen

Description automatically generated

1. Dùng Flat file source để làm nguồn dữ liệu:

A close-up of a sign

Description automatically generated

1. Sử dụng Lookup Transformation để tìm kiếm các dữ liệu liên quan:
   1. Thiết lập Lookup Transformation

*Ví dụ với bảng MaturitySize\_Dim*

A screenshot of a computer

Description automatically generated

Thiết lập No Cache và kết nối với dữ liệu qua OLD DB Connection manager.

Ignore các dữ liệu không match với thực thể

A screenshot of a computer

Description automatically generated

Kết nối với với cơ sở dữ liệu và chọn bảng MaturitySize\_Dim

A screenshot of a computer

Description automatically generatedLookup các dữ liệu có match với Maturity Size trong bảng MaturitySize\_Dim

Và chọn sử dụng các trường MaturitySizeID và MaturitySize

A screen shot of a computer

Description automatically generated

Sử dụng output match với cài đặt Lookup Transformation ở trên để kết nối với hành động tiếp theo

Tương tự, ta cài đặt Lookup Transformation cho các trường dữ liệu khác của bảng khác để được sơ đồ như hình dưới

A diagram of a computer

Description automatically generated with medium confidence

1. A screen shot of a computer

   Description automatically generatedNhập dữ liệu vào Fact Table

Sử dụng công cụ OLD DB Destination để nhập dữ liệu sau các Lookup Transformation

A screenshot of a computer

Description automatically generatedKết nối với bảng Pet\_Fact

A screenshot of a computer

Description automatically generated

Mapping các dữ liệu có được từ output alias của các Lookup Transformation phù hợp với các cột của Fact Table

Sau khi hoàn thành, ta được Data Flow của Fact table như hình dưới:

A diagram of a diagram

Description automatically generated with medium confidence

1. **Viết SQL để thực hiện các truy vấn nghiệp vụ**
2. Số lượng thú nuôi chưa được tiêm vaccine và chưa được sổ giun, phân theo từng khu vực (state)

A computer screen shot of a number

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Số lượng từng loài thuần chủng (Không được lai giữa 2 loại khác nhau)

A computer screen shot of a code

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Phí nhận nuôi theo trung bình theo kích thước trưởng thành

A white background with black text

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Phí nhận nuôi trung bình theo giống

A white background with black text

Description automatically generated

A screenshot of a computer

Description automatically generated

1. **Thiết kế các ETL chạy song song**

Các ETL cho việc nhập dữ liệu vào các Dimension table được nhóm chung vào Sequence Contatiner để các ETL chạy song song.

Sau khi các ETL này hoàn thành thì việc nhập dữ liệu vào Fact table mới được thực thi.

A screenshot of a computer

Description automatically generated