1. **Thiết kế ERD**

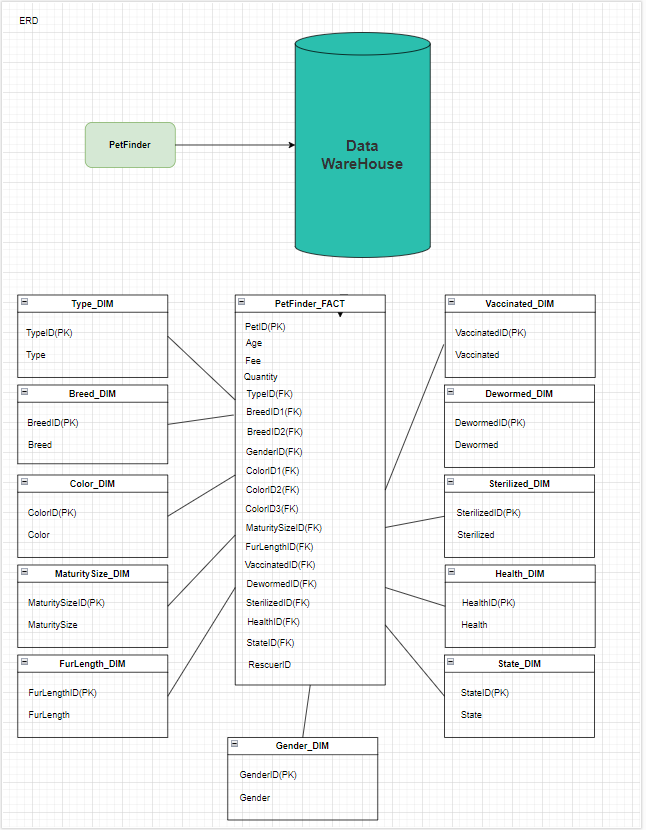
Sau khi tải và có 1 số phân tích tổng quan ban đầu file Dataset chúng ta xác định tất cả tập dữ liệu là thông tin của bảng Fact vì có chứa các quan sát hoặc dữ liệu đo lường ví dụ cột Quantity và các thông tin lặp lại nhiều lần ví dụ như Type, Gender, Color…

Dựa vào bảng Fact chúng ta có thể thiết kế một số các bảng DIM chứa các thông tin chi tiết của bảng Fact và Dim kết nối với Fact thông qua các khóa, cụ thể khóa chính của các bảng Dim sẽ là khóa ngoại của bảng Fact. Dữ liệu trên bảng Dim sẽ là những dữ liệu duy nhất(không lặp lại) lấy từ các cột có giá trị lặp lại trong bảng Fact, trường hợp cột của bảng Fact chứa các dữ liệu duy nhất thì sẽ giữ lại bảng Fact. Mục đích của Dim là để lọc và nhóm dữ liệu trong bảng Fact, phục vụ các yêu cầu nghiệp vụ về phân tích dữ liệu.

Tìm kiếm và xác định các bảng DIM, FACT dựa theo Dataset:

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Tên cột** | **Chú thích** | **Loại Bảng** |
| 1 | Type | Có thể nhóm được theo đặc trưng | DIM |
| 2 | Breed1  Breed1 | Có thể nhóm được theo đặc trưng | DIM |
| 3 | Color1  Color2  Color3 | Có thể nhóm được theo đặc trưng | DIM |
| 4 | MaturitySize | Có thể nhóm được theo đặc trưng | DIM |
| 5 | FurLength | Có thể nhóm được theo đặc trưng | DIM |
| 6 | Gender | Có thể nhóm được theo đặc trưng | DIM |
| 7 | Vaccinated | Có thể nhóm được theo đặc trưng | DIM |
| 8 | Dewormed | Có thể nhóm được theo đặc trưng | DIM |
| 9 | Sterilized | Có thể nhóm được theo đặc trưng | DIM |
| 10 | Health | Có thể nhóm được theo đặc trưng | DIM |
| 11 | State | Có thể nhóm được theo đặc trưng | DIM |
| 12 | PetID  Age  Quantity  Fee  RescuerID | Dữ liệu là duy nhất hoặc các dữ liệu đo lường. | FACT |

Chúng ta có ERD như sau:



**Tổng hợp lại ta có các bảng sau:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Tên bảng** | **Cột** | **Khóa chính** | **Khóa ngoại** |
| 1 | Type\_DIM | TypeID, Type | TypeID |  |
| 2 | Breed\_DIM | BreedID, Breed | Breed ID |  |
| 3 | Color\_DIM | ColorID, Color | ColorID |  |
| 4 | MaturitySize\_DIM | MaturitySizeID, MaturitySize | MaturitySizeID |  |
| 5 | FurLength\_DIM | FurLengthID, FurLength | FurLengthID |  |
| 6 | Gender\_DIM | GenderID, Gender | GenderID |  |
| 7 | Vaccinated\_DIM | VaccinatedID, Vaccinated | VaccinatedID |  |
| 8 | Dewormed\_DIM | DewormedID, Dewormed | DewormedID |  |
| 9 | Sterilized\_DIM | SterilizedID, Sterilized | SterilizedID |  |
| 10 | Health\_DIM | HealthID, Health | HealthID |  |
| 11 | State\_DIM | StateID, State | StateID |  |
| 12 | PetFinder\_FACT | PetID, Age, Quantity, Fee, RescuerID, TypeID,  BreedID1, BreedID2,  ColorID1,  ColorID2,  ColorID3,  MaturitySizeID,  FurLengthID,  GenderID,  VaccinatedID,  DewormedID,  SterilizedID,  HealthID,  StateID | PetID | TypeID,  BreedID1, BreedID2,  ColorID1,  ColorID2,  ColorID3,  MaturitySizeID,  FurLengthID,  GenderID,  VaccinatedID,  DewormedID,  SterilizedID,  HealthID,  StateID |
| 13 | PetFinder\_Staging | PetID, Age, Quantity, Fee, Type,  Breed1, Breed2,  Color1,  Color2,  Color3,  MaturitySize,  FurLength,  Gender,  Vaccinated,  Dewormed,  Sterilized,  Health,  State,  RescuerID,  TypeID,  BreedID1,  BreedID2,  ColorID1,  ColorID2,  ColorID3,  MaturitySizeID,  FurLengthID,  GenderID,  VaccinatedID,  DewormedID,  SterilizedID,  HealthID,  StateID | PetID |  |

1. **Xác định các truy vấn nghiệp vụ**

* Tổng cộng có bao nhiêu vật nuôi đã được tiêm phòng tính theo kích cỡ khi trưởng thành?
* Có bao nhiêu con chó đã được triệt sản? Chiếm tỷ lệ bao nhiêu phần trăm trên tổng số chó?
* Có bao nhiêu con mèo lông dài?
* Tổng cộng có bao nhiêu vật nuôi trong từng bang?

1. **Thiết kế Database và Xây dựng ETL**

* Tạo database PetFinderDataSource và bảng PetFinder chứa dữ liệu nguồn.

File : [SQLQuery\_PetFinderDataSource](file:///F:\Funix\DAP303x_01-A_VN%20Truy%20vấn%20nâng%20cao%20và%20xây%20dựng%20kho%20dữ%20liệu\SQLQuery_PetFinderDataSource.sql)

* Tạo database PetFinder chứa dữ liệu Destination là các bảng, DIM, FACT

Tạo thêm 1 bảng PetFinder\_Staging chứa tất cả 19 cột của dữ liệu nguồn ngoài ra còn có các cột là khóa ngoại của bảng FACT tham chiếu tới khóa chính bảng DIM.

Chức năng của PetFinder\_Staging là 1 bảng nguồn trung gian, chúng ta sẽ sử dụng PetFinder\_Staging là nguồn để ETL cho các DIM\_TABLE, sau khi đã ETL được các bảng DIM thì khi đó bảng DIM sẽ có các khóa chính, khi đó sẽ update ngược khóa chính từ các bảng DIM vào lại PetFinder\_Staging. Khi PetFinder\_Staging đã có đầy đủ các khóa chính thì chúng ta sẽ ETL cho FACT thông qua PetFinder\_Staging.

File : [SQLQuery\_Creat\_DIM\_FACT\_TABLE](file:///F:\Funix\DAP303x_01-A_VN%20Truy%20vấn%20nâng%20cao%20và%20xây%20dựng%20kho%20dữ%20liệu\SQLQuery_Creat_DIM_FACT_TABLE.sql)

* Khởi động SSIS ETL cho bảng PetFinder\_Staging từ dữ liệu nguồn là bảng PetFinder thuộc database PetFinderDataSource.
* ETL lần lượt cho tất cả các DIM\_Table sau đó update lại các cột là khóa chính của các bảng DIM cho các cột tương ứng của bảng PetFinder\_Staging.
* ETL cho FACT\_Table từ PetFinder\_Staging.

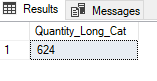
ETL File: [ASM2.sln](file:///F:\Funix\DAP303x_01-A_VN%20Truy%20vấn%20nâng%20cao%20và%20xây%20dựng%20kho%20dữ%20liệu\DAP303x_asm2_tuyenttFX18918@funix.edu.vn\ASM2.sln)

1. **Kết quả truy vấn nghiệp vụ:**

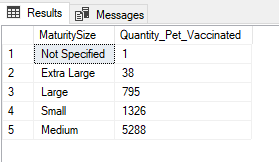
* Tổng cộng có bao nhiêu vật nuôi trong từng bang?



* Có bao nhiêu con mèo lông dài?



* Tổng cộng có bao nhiêu vật nuôi đã được tiêm phòng tính theo kích cỡ khi trưởng thành?



* Có bao nhiêu con chó đã được triệt sản? Chiếm tỷ lệ bao nhiêu phần trăm trên tổng số chó

