Lời cảm ơn

Đầu tiên, nhóm chúng em xin gửi lời cảm ơn đến Thầy – Th.S Nguyễn Văn Thành – Giảng viên phụ trách môn Kho Dữ Liệu – Trường Đại học Sư phạm Kỹ thuật TP. Hồ Chí Minh vì những đóng góp của thầy trong thời gian qua. Thầy đã cung cấp những kiến thức, hướng dẫn, đóng góp, xây dựng trong quá trình học để nhóm em có thể hoàn thành đồ án môn học một cách tốt nhất.

Kho dữ liệu là một trong những kiến thức đang ngày càng phổ biến cùng với OLAP, nhóm chúng em đã lên ý tưởng thực hiện xây dựng một kho dữ liệu cùng với các quy trình nghiệp vụ. Nhóm chúng em quyết định xây dựng đồ án “Xây dựng và khai thác kho dữ liệu về lĩnh vực hàng không”. Trong quá trình thực hiện đồ án, vận dụng những kiến thức học được từ lý thuyết đến thực hành trong quá trình học tập, cùng với sự tìm hiểu các kiến thức, công cụ mới, nhóm chúng em đã cố gắng hoàn thiện đồ án một cách tốt nhất.

Nếu có những chỗ chưa hoàn thiện, nhóm chúng em mong nhận được sự đóng góp từ thầy nhằm cải thiện và coi đó là một kinh nghiệm để có thể hoàn thành tốt hơn.

Nhóm chúng em một lần nữa xin chân thành cảm ơn thầy!

Lời Nhận Xét Của Giảng Viên

…………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………….………………………………………………………………………………..………

Chương 1: Lý do chọn đề tài và giới thiệu tổng quan Dataset

* 1. Lý do hình thành đồ án

1.1.1 Vấn đề đặt ra

Ngành hàng không là một trong những ngành trọng điểm đối với thế giới cũng như Việt Nam. Trải qua lịch sử hơn 100 năm đầy biến động, ngành hàng không đã có những biến động cả về công nghệ lẫn dịch vụ.

Từ năm 1994 đến năm 2002 nổi lên với sự phát triển của mô hình hàng không giá rẻ, có ý nghĩa quan trọng với ngành hàng không ngày nay. Các công ty con và trung tâm phân phối được mở rộng toàn thế giới cùng với sự phát triển của thương mại điện tử cho các hệ thống bán vé máy bay trực tuyến.

Tháng 12/2019, tâm dịch Covid-19 nhanh chóng lan rộng ra toàn thế giới khiến tất cả các nước phải áp dụng các biện pháp hạn chế di chuyển, điểu này đặc biệt nghiêm trọng cho ngành hàng không thế giới. American Airlines cắt giảm 75% số chuyến bay quốc tế, Korean Air giảm 80% lịch trình,… tóm lại, hầu như các chuyến bay trên thế giới đều bị cắt giảm đáng kể.

Trong 2 năm vừa qua, ngành hàng không quốc tế cũng như Việt Nam đang có dấu hiệu phục hồi trở lại

Các kiến thức về ngành hàng không rất đa dạng như thông tin về các hãng hàng không trên toàn thế giới, các sân bay, chuyến bay,… và phải kể đến quy trình đặt vé máy bay hay còn gọi là booking. Những data kể trên rất phù hợp để xây dựng một kho dữ liệu về lịch trình chuyến bay cũng như booking.

Vì vậy, nhóm chúng em quyết định xây dựng một kho dữ liệu về các chuyến bay và đặt vé máy bay của ngành hàng không thế giới

<https://thinksmart.com.vn/hon-100-nam-ra-doi-va-phat-trien-nganh-hang-khong/>

1.1.2 Giải pháp

Dựa trên nhu cầu phân tích, thống kê và khai thác dữ liệu về ngành hàng không thế giới. Giải pháp được đưa ra là xây dựng một kho dữ liệu nhằm mục đích khai thác, phân tích và tạo báo cáo tổng về các chuyến bay và booking.

1.1.3 Mục tiêu và ý nghĩa dự án

Đầu tiêu là xây dựng một kho dữ liệu và OLAP

Data Warehouse hay kho dữ liệu lưu trữ data từ nhiều nguồn, giúp tăng cường hiệu suất báo cáo và phân tích. Dựa vào điều đó cũng như các kiến thức đã học được trong quá trình học tập, nhóm chúng em quyết định xây dựng một kho dữ liệu OLAP để thấy được hiệu quả của nó.

Thứ hai là khai thác và phân tích data về hàng không trong thực tế. Khai thác các thông tin của hàng không quốc tế nhằm có cái nhìn chung về kiến thức cũng như quy trình vận hành cơ bản của một hệ thống đặt vé máy bay.

* 1. Giới thiệu tổng quan dataset

1.2.1 Nguồn dữ liệu sử dụng

Nguồn dataset được lấy từ github.com/fortunewalla/airportdb

1.2.2 Mô tả chi tiết dữ liệu

1.2.2.1 Tổng quát

Dataset có tổng cộng 12 tables với hơn 450.000 dòng mô tả các hoạt động của các chuyến bay trên toàn thế giới vào năm 2015, mô tả hoạt động đặt vé máy bay cùng với một tập khách hàng, mô tả các sân bay, hãng hàng không, máy bay trên toàn thế giới, ngoài ra dataset còn thống kê các nhân viên làm việc cho các phòng ban khác nhau cùng với mức lương của họ.

1.2.2.2 Mô tả các bảng dataset

|  |  |
| --- | --- |
| Tên bảng | Mô tả |
| airport | Mô tả các sân bay theo Hiệp hội Vận tải Hàng không Quốc tế IATA và theo Tổ chức Hàng không Dân dụng Quốc tế ICAO |
| airport\_geo | Mô tả chi tiết thành phố, đất nước, vị trí các sân bay |
| airline | Mô tả các hãng hàng không theo Hiệp hội Vận tải Hàng không Quốc tế IATA |
| airplane | Mô tả các máy bay thuộc các hãng hàng không |
| airplane\_type | Mô tả các kiểu máy bay |
| flight | Mô tả hoạt động các chuyến bay bao gồm số hiệu chuyến bay, nơi đi, nơi đến, giờ cất cánh, hạ cánh,… |
| flight\_schedule | Mô tả lịch trình bay của các chuyến bay |
| booking | Mô tả việc đặt vé máy bay của các khách hàng |
| passenger | Mô tả khách hàng |
| passenger\_detail | Mô tả chi tiết thông tin khách hàng |
| employee | Mô tả thông tin nhân viên |

1.2.3 Các công cụ được sử dụng

1.2.3.1 Visual Studio 2022

Visual studio là một phần mềm hỗ trợ đắc lực hỗ trợ công việc lập trình website. Công cụ này được tạo lên và thuộc quyền sở hữu của ông lớn công nghệ Microsoft. Năm 1997, phần mềm lập trình nay có tên mã Project Boston. Nhưng sau đó, Microsoft đã kết hợp các công cụ phát triển, đóng gói thành sản phẩm duy nhất.

Phần mềm lập trình Visual Studio của Microsoft hỗ trợ sử dụng trên nhiều nền tảng khác nhau. Không giống như các trình viết code khác, Visual Studio sử dụng được trên cả Windows, Linux và Mac Systems. Điều này cực kỳ tiện lợi cho lập trình viên trong quá trình ứng dụng.

Không chỉ hỗ trợ đa nền tảng, Visual Studio cũng cho phép sử dụng nhiều ngôn ngữ lập trình khác nhau từ C#, F#, C/C++, HTML, CSS, Visual Basic, JavaScript,… Bởi vậy, Visual Studio có thể dễ dàng phát hiện và thông báo cho bạn khi các chương trình có lỗi.

Mặc dù Visual Studio có hệ thống các ngôn ngữ hỗ trợ lập trình khá đa dạng. Nhưng nếu lập trình viên muốn sử dụng một ngôn ngữ khác, bạn có thể dễ dàng tải xuống các tiện ích mở rộng. Tính năng hấp dẫn này được hoạt động như một phần chương trình độc lập nên không lo làm giảm hiệu năng của phần mềm.

Phần lớn các tệp dữ liệu đoạn mã của Visual Studio đều được đặt trong các thư mục tương tự nhau. Đồng thời, Visual Studio cũng cung cấp một số thư một cho các tệp đặc biệt để bạn lưu trữ an toàn, dễ tìm, dễ sử dụng hơn.

Với Visual Studio, bạn có thể hoàn toàn yên tâm về tính lưu trữ, bởi phần mềm đã được kết nối GIT và một số kho lưu trữ an toàn được sử dụng phổ biến hiện nay.

Visual Studio sở hữu tính năng màn hình đa nhiệm, cho phép người dùng mở cùng lúc nhiều tập tin, thư mục dù chúng có thể không liên quan tới nhau.

Khi sử dụng code vào trong lập trình, với Visual Studio, công cụ này có thể đề xuất tới các lập trình viên một số tùy chọn thay thế nhằm điều chỉnh đôi chút để đoạn code áp dụng thuận tiện hơn cho người dùng.

Phần mềm Visual Studio cũng tích hợp các loại thiết bị đầu cuối, giúp người dùng không cần chuyển đổi giữa hai màn hình hay trở về thư mục gốc khi thực hiện một thao tác cần thiết nào đó.

https://visualcpp.net/visual-studio-la-gi-gioi-thieu-phan-mem-visual-studio/

Cho đến nay, nhà phát triển Microsoft vẫn đang không ngừng cải tiến và ứng dụng các công nghệ mới cho Visual Studio. Bởi vậy, đây là “cánh tay đắc lực” hỗ trợ cho các developers, data administrators, testers và bất kỳ ai có nhu cầu tiếp cận, tìm hiểu về lập trình đều có thể sử dụng Visual Studio.

1.2.3.2 SQL Server 18

1.2.3.3

Chương 2: Thiết kế kho dữ liệu dùng kỹ thuật mô hình hóa dữ liệu theo chiều (Dimensional Modeling)

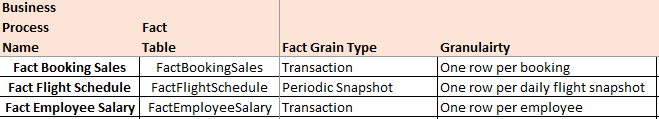
2.1 Thiết kế kho dữ liệu mức cao (High Level Design)

2.1.1 Xây dựng Business Process

Các yêu cầu business đặt ra:

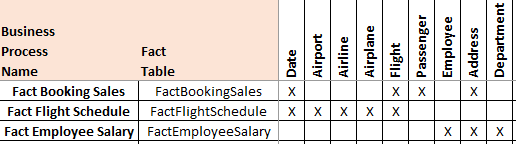
* Báo cáo doanh số đặt vé máy bay (Booking Sales Reporting): theo dõi doanh số bán vé máy bay (booking) dựa trên khách hàng (passenger), chuyến bay (flight) để biết được rằng khách hàng nào đặt vé nhiều nhất, chuyến bay nào được đặt nhiều nhất.
* Lịch trình chuyến bay (Flight Schedule Monitoring): ghi lại hoạt động của các chuyến bay(flight) theo ngày để biết được rằng chuyến bay bay nhiều nhất, hãng hàng không (airline) nào có nhiều chuyến bay nhất, máy bay (airplane) nào bay nhiều nhất, điểm đến (airport) nhiều nhất.
* Phân tích lương nhân viên (Employee Salary Analysis): thống kê lương nhân viên (employee) theo tháng làm việc tại ngành hàng không để biết được rằng nhân viên thuộc phòng ban (department) nào được trả lương cao nhất.

2.1.2 Xác định mức độ chi tiết (Grain) cho các business process



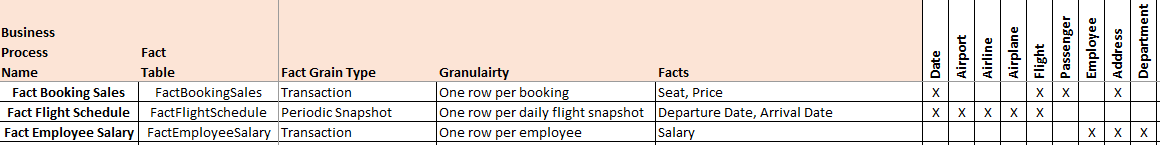
* Hoạt động Booking Sales Reporting gắn với Fact Booking Sales ghi lại mỗi dòng thể hiện việc bán một vé máy bay, thuộc loại Transaction.
* Hoạt động Flight Schedule Monitoring gắn với Fact Flight Schedule ghi mỗi dòng cho hoạt động bay của một chuyến bay, thuộc loại Periodic Snapshot.
* Hoạt động Employee Salary Analysis gắn với Fact Employee Salary ghi lại mỗi dòng với mỗi lương của một nhân viên, thuộc loại Transaction.

2.1.3 Xác định các Dimension



* Fact Booking Sales liên quan đến bảng Flight, Passenger, phân cấp khách hàng theo bảng Address, bảng Date.
* Fact Flight Schedule liên quan đến bảng Airport, Airline, Airplane, Flight và Date.
* Fact Employee Salary liên quan đến bảng Employee, phân cấp Employee theo Address, Department.

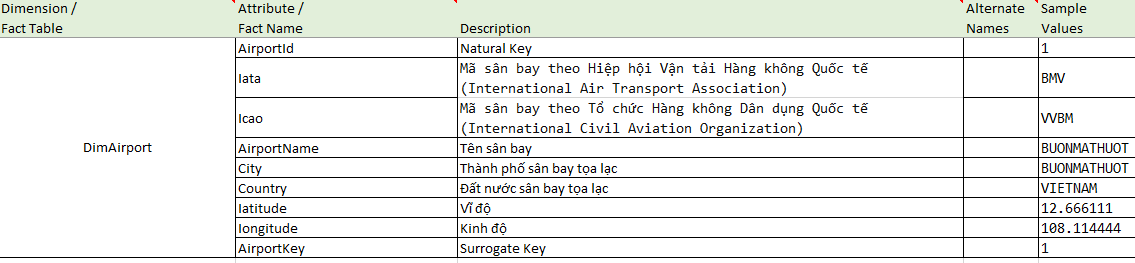
2.1.4 Xác định các Fact Measure



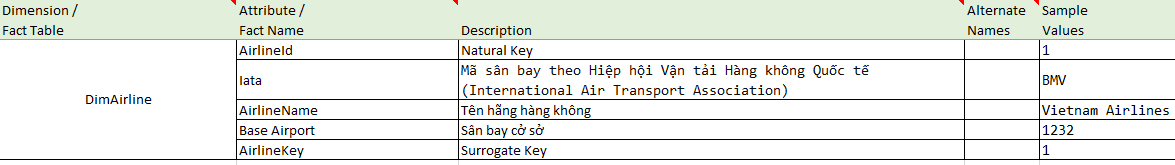
* Fact Booking Sales ghi lại Seat (ghế ngồi) và Price (giá vé theo USD) của một khách hàng khi đặt vé.
* Fact Flight Schedule ghi lại Departure Date (ngày bay) và Arrival Date (ngày hạ cánh) của một chuyến bay
* Fact Employee Salary ghi lại Salary (lương theo tháng) ứng với mỗi nhân viên

2.2 Thiết kế kho dữ liệu mức chi tiết (Detailed Level Design)

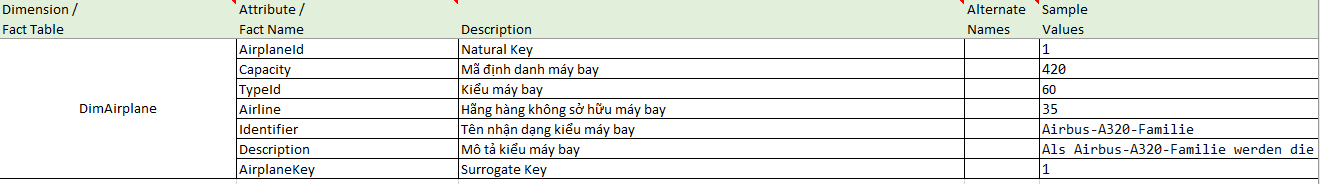
2.2.1 Mô tả các bảng Dimension



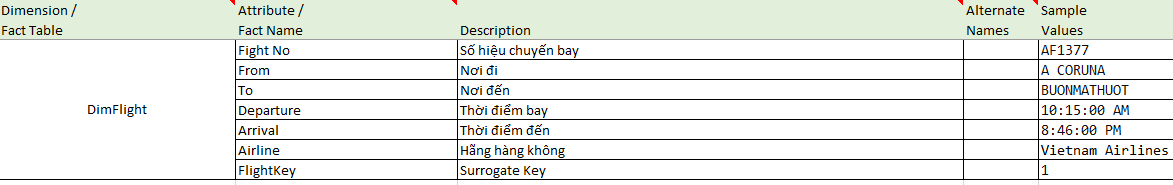
* Bảng DimAirport gồm các trường AirportId, Iata, Icao, AirportName, City, Country, Iatitude, Iongitude, AirportKey với mô tả và giá trị ví dụ



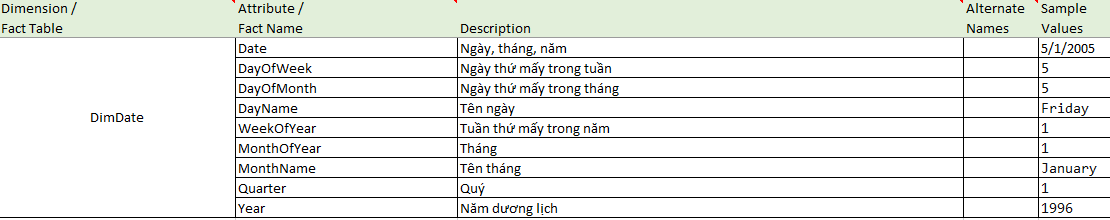
* Bảng DimAirline gồm các trường AirlineId, Iata, AirlineName, BaseAirport, AirlineKey với mô tả và giá trị ví dụ



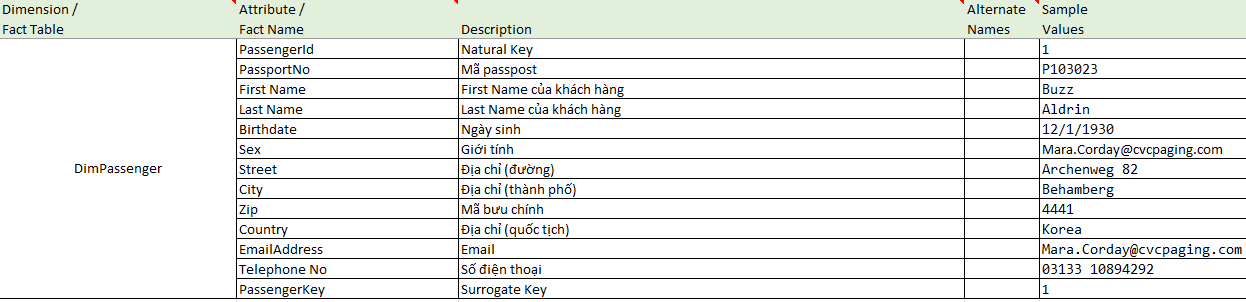
* Bảng DimAirplane gồm các trường AirplaneId, Capacity, TypeId, Airline, Identifier, Description, AirplaneKey với mô tả và giá trị ví dụ



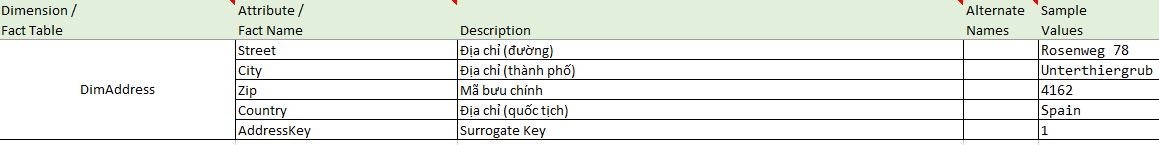
* Bảng DimFlight gồm các trường FlightNo, From, To, Departure, Arrival, Airpline, FlightKey với mô tả và giá trị ví dụ



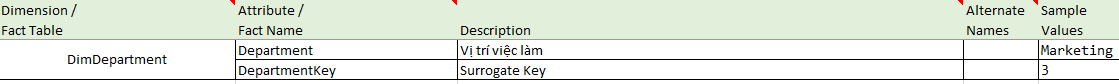
* Bảng DimDate mô tả một dòng là một ngày



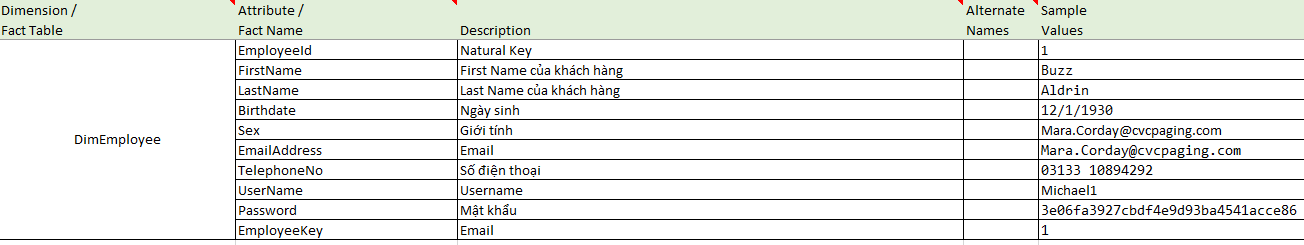
* Bảng DimPassenger gồm các trường PassengerId, PassportNo, FirstName, LastName, Birthdate, Sex, Street, City, Zip, Country, EmailAddress, TelephoneNo, PassengerKey với mô tả và giá trị ví dụ



* Bảng DimAddress gồm các trường Street, City, Zip, Country, AddressKey với mô tả và giá trị ví dụ

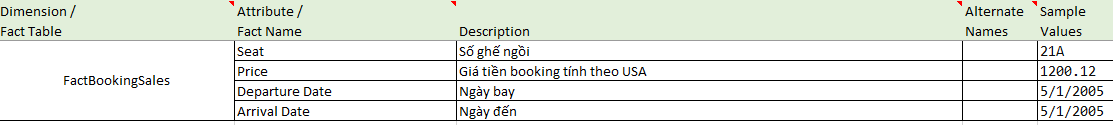


* Bảng DimDepartment gồm các trường Department, DepartmentKey với mô tả và giá trị ví dụ

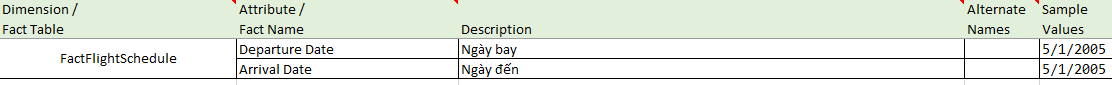


* Bảng DimEmployee gồm các trường EmployeeId, FirstName, LastName, BirthDate, Sex, EmailAddress, TelephoneNo, UserName, Password, EmployeeKey với mô tả và giá trị ví dụ

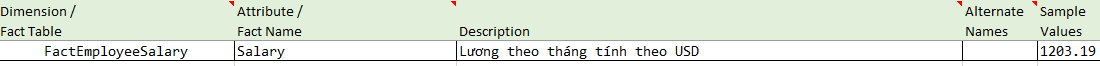
2.2.2 Mô tả các bảng Fact



* FactBookingSales ghi lại Seat, Price, DepartureDate, ArrivalDate cho mỗi booking



* FactFlightSchedule ghi lại DepartureDate và ArrivalDate cho mỗi chuyến bay



* FactEmployeeSalary ghi lại Salary cho mỗi nhân viên