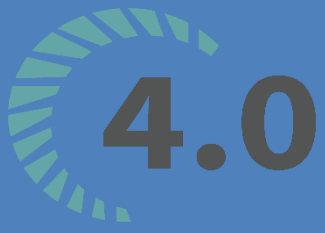
BỘ MÔN HỆ THỐNG THÔNG TIN – KHOA CÔNG NGHỆ THÔNG TIN
ĐẠI HỌC KHOA HỌC TỰ NHIÊN THÀNH PHỐ HỒ CHÍ MINH, ĐẠI HỌC QUỐC GIA TP HCM

HỆ THỐNG THÔNG TIN PHỤC VỤ TRÍ TUỆ KINH DOANH



Nhóm: TTKD-23

GV phụ trách: Hồ Thị Hoàng Vy

ĐỒ ÁN MÔN HỌC - HTTT Phục Vụ Trí Tuệ Kinh Doanh

HỌC KỲ I– NĂM HỌC 2021-2022

KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
227 Nguyễn Văn Cừ, Phường 4, Quận 5, TP.HCM
Điện Thoại: (08) 38.354.266 - Fax:(08) 38.350.096

BẢNG THÔNG TIN CHI TIẾT NHÓM

MSSV	Họ tên	Email
1712849	Tô Hoàng Trung	1712849@student.hcmus.edu.vn
1712526	Nguyễn Quốc Khánh	1712526@student.hcmus.edu.vn
18120217	Nguyễn Trần Ái Nguyên	18120217@student.hcmus.edu.vn
18120175	Nguyễn Vũ Hà	18120175@student.hcmus.edu.vn

ĐÁNH GIÁ HOÀN THÀNH CÔNG VIỆC		
Người thực hiện	**Công việc**	**Mức độ hoàn thành**
Tô Hoàng Trung	Postcodes Viết báo cáo	100%
Nguyễn Quốc Khánh	Vehicles 1114	100%
Nguyễn Trần Ái Nguyên	Casualties 1114 Accidents 1114 Viết Báo cáo	100%
Nguyễn Vũ Hà	PCD_OA_LSOA_MSOA_LAD_AUG21_UK_LU	100%

KẾT QUẢ

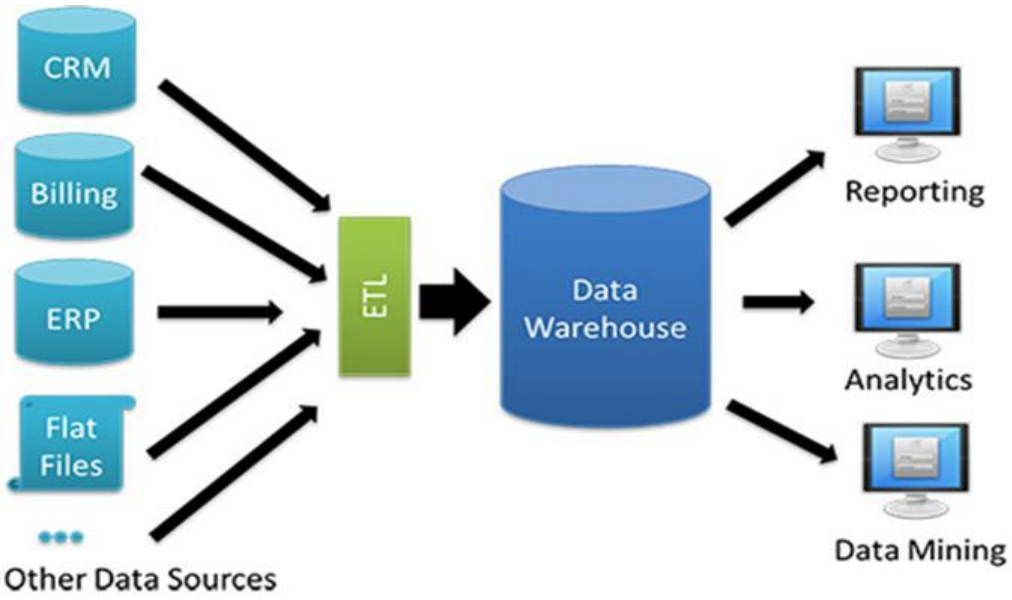
Contents

1. PHÂN TÍCH

1.1 Giới thiệu

1.1.1 Mô tả đồ án

Từ kiến thức lý thuyết đã học về KDL, OLAP, ETL, mining…, đồ án **Xây dựng và phân tích Dữ liệu về lộ trình Yellow Taxi ở New York City (NYC) từ năm 2014 đến 2018** giúp sinh viên thực hành xây dựng một KDL cụ thể, biết triển khai ETL để rút trích dữ liệu từ nhiều nguồn, biết cách khai thác KDL với report, OLAP, mining, tạo job định kì thực hiện ETL.



Hình ảnh tổng quan kiến trúc của kho dữ liệu(Data warehouse)

1.1.2 Mục tiêu đồ án

Đồ án này nhằm mục tiêu đạt được các chuẩn đầu ra sau:

1. Thiết kế được lược đồ chuẩn hóa, đa chiều(sao, bông tuyết) dựa vào dữ liệu hệ thống tác vụ và yêu cầu phân tích từ tình huống cho trước.
2. Triển khai quy trình ETL để rút trích dữ liệu từ nhiều nguồn, biến đổi, làm sạch dữ liệu, nạp dữ liệu vào kho dữ liệu(KDL) sử dụng SSIS.
3. Xây dựng KDL đa chiều sử dụng SSAS và giải thích được lựa chọn phép toán OLAP phù hợp đối với 1 số yêu cầu phân tích.
4. Sử dụng 1 số công cụ biểu diễn dữ liệu(SSRS, powerBI, exel…) để biểu diễn kết quả phân tích, khai thác được(report, dashboard)
5. Sử dụng SSAS và áp dụng các kĩ thuật mining tích hợp để thực hiện khai thác dữ liệu từ KDL xây dựng được.

1.2 Yêu cầu chức năng

1.2.1 Usecase các chức năng

1.2.2 Đặc tả chức năng

1.3 Yêu cầu phi chức năng

1.4 Yêu cầu dữ liệu và chất lượng dữ liệu

1.4.1 Nguồn Dữ liệu

STT	Tên nguồn dữ liệu	Mô tả nguồn dữ liệu
1	Postcodes	Danh sách mã bưu chính tại nước Anh
2	PCD_OA_LSOA_MSOA_LAD_AUG21_UK_LU	Mô tả chi tiết Postcodes
3	Vehicles 1114	Danh sách thông tin phương tiện
4	Casualties 1114	Danh sách thông tin thương vong
5	Accidents 1114	Danh sách thông tin tai nạn

1.4.2 Mô tả ý nghĩa thuộc tính nguồn dữ liệu

1.4.2.1 Thuộc tính nguồn Postcodes

STT	Column	Description
1	postcode	Mã bưu chính
2	easting	Độ Đông
3	northing	Độ Bắc
4	latitude	Vĩ độ
5	longitude	Kinh độ
6	city	Thành phố
7	county	Quốc gia
8	country_code	Mã quốc gia
9	country_name	Tên quốc gia
10	iso3166-2	Mã địa lý
11	region_code	Mã vùng
12	region_name	Tên vùng

1.4.2.2 Thuộc tính nguồn PCD_OA_LSOA_MSOA_LAD_AUG21_UK_LU

STT	Column	Description
1	pcd7	postcode 7 kí tự
2	pcd8	postcode 8 kí tự
3	Pcds	Postcode có khoảng trắng ngăn cách giữa phần quận và phần đơn vị ngành
4	dointr	Ngày xuất hiện gần nhất của ngày giới thiệu mã bưu điện
5	doterm	Lần xuất hiện gần nhất của ngày chấm dứt mã bưu điện. YYYYMM (year and month)
6	usertype	Cho biết postcode là của người dùng nhỏ hay lớn
7	oa11cd	Mã của Output Area được định nghĩa vào năm 2011
8	lsoa11cd	Mã của Lower Layer Super Output Areas được định nghĩa vào năm 2011
9	msoa11cd	Mã của Middle Layer Super Output Areas được định nghĩa vào năm 2011
10	ladcd	Mã của chính quyền địa phương

1.4.2.3 Thuộc tính nguồn Vehicles 1114

STT	Column	Description
1	Accident_Index	Chỉ số tai nạn
2	Vehicle_Reference	Tham khảo phương tiện
3	Vehicle_Type	Loại phương tiện
4	Vehicle_Manoeuvre	Phương tiện di chuyển
5	Hit_Object_in_Carriageway	Đối tượng/vật thể va chạm trong đường di chuyển
6	1st_Point_of_Impact	Điểm tác động đầu tiên
7	Was_Vehicle_Left_Hand_Drive?	Có phải xe tay lái bên trái không?

8	Sex_of_Driver	Giới tính của người lái xe
9	Age_of_Driver	Tuổi của lái xe
10	Age_Band_of_Driver	Nhóm tuổi của người lái xe
11	Engine_Capacity_(CC)	Công suất động cơ(CC)
12	Propulsion_Code	Mã lực đẩy
13	Age_of_Vehicle	Tuổi xe
14	Driver_IMD_Decile	Trình điều khiển IMD Decile

1.4.2.4 Thuộc tính nguồn Casualties 1114

STT	Column	Description
1	Accident_Index	Chỉ số tai nạn
2	Vehicle_Reference	ID phương tiện
3	Casualty_Reference	ID nạn nhân
4	Casualty_Class	Người thương vong(tài xế, khách, đi bộ)
5	Sex_of_Casualty	Giới tính của nạn nhân
6	Age_of_Casualty	Tuổi tai nạn
7	Age_Band_of_Casualty	Nhóm tuổi của nạn nhân
8	Casualty_Severity	Mức độ thương vong nghiêm trọng
9	Casualty_Type	Loại tai nạn
10	Casualty_Home_Area_Type	Vùng xảy ra tai nạn

1.4.2.5 Thuộc tính nguồn Accidents 1114

STT	Column	Description
1	Accident_Index	Chỉ số tai nạn
2	Location_Easting_OSGR	Vị trí Easting OSGR (Không có nếu không biết)
3	Location_Northing_OSGR	Vị trí Northing OSGR (Không có nếu không biết)
4	Longitude	Kinh độ (Null nếu không xác định)
5	Latitude	Vĩ độ (Null nếu không biết)

6	Accident_Severity	Mức độ nghiêm trọng của tai nạn
7	Number_of_Vehicles	Số lượng phương tiện trong vụ tai nạn
8	Number_of_Casualties	Số lượng thương vong
9	Date	Ngày (DD / MM / YYYY) xảy ra tai nạn
10	Day_of_Week	Ngày trong tuần xảy ra tai nạn
11	Time	Thời gian (HH: MM) xảy ra tai nạn
12	Local_Authority_(District)	Mã quận xảy ra tai nạn
13	Local_Authority_(Highway)	Mã xa lộ xảy ra tai nạn
14	Road_Type	Loại đường
15	Speed_limit	Tốc độ giới hạn
16	Light_Conditions	Điều kiện ánh sáng
17	Weather_Conditions	Điều kiện thời tiết
18	Road_Surface_Conditions	Điều kiện mặt đường
19	Special_Conditions_at_Site	Các tác động ngoại cảnh đặc biệt
20	Urban_or_Rural_Area	Khu vực thành thị hoặc nông thôn
21	LSOA_of_Accident_Location	Chỉ số dùng để cải thiện việc báo cáo thống kê 1 diện tích nhỏ ở Anh và Wales

1.4.3 Data Cleansing

2. KIẾN TRÚC DATASTORE

3. THIẾT KẾ

3.1 DATA STORE: Stage

Data Store: Stage Data Store

3.1.1 COLUMNS

3.1.1.1 Thuộc tính bảng TAXITRIP_2014_Stage

Column	Key Type	Data Type	Is Nullable	Default Value	Description

3.1.2 CONSTRAINTS

3.1.3 DATASTORE LOGICAL MODEL

3.2 DATA STORE: NDS

Data Store: Normalization Data Store

Schema Type: Entity-Relationship 3NF

3.2.1 COLUMNS

3.2.1.1 Thuộc tính bảng

Column	Key Type	Data Type	Is Nullable	Default Value	Description

3.2.2 CONSTRAINTS

3.2.2.1 Ràng buộc bảng

Constraint Name	Constraint Type	Table	Column	Description

3.2.3 DATA STORE LOGICAL MODEL