

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG

VŨ ĐỨC THẢO

**GIẢI PHÁP KHO DỮ LIỆU THỜI GIAN
THỰC CHO HỆ THỐNG SIÊU THỊ**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Thái Nguyên – 2014

**ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

VŨ ĐỨC THẢO

**GIẢI PHÁP KHO DỮ LIỆU THỜI GIAN
THỰC CHO HỆ THỐNG SIÊU THỊ**

Chuyên ngành: Khoa học máy tính

Mã số: 60 48 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC:

TS. HOÀNG ĐỖ THANH TÙNG

Thái Nguyên – 2014

LỜI CẢM ƠN

Trước hết, tôi xin bày tỏ lòng biết ơn sâu sắc tới TS. Hoàng Đỗ Thanh Tùng đã tận tình hướng dẫn, chỉ bảo và cung cấp những tài liệu rất hữu ích để tôi có thể hoàn thành luận văn.

Xin cảm ơn lãnh đạo Trường Đại học Công nghệ Thông tin và Truyền thông - Đại học Thái Nguyên, Trường Cao đẳng Dược Phú Thọ đã tạo điều kiện giúp đỡ tôi về mọi mặt trong suốt quá trình thực hiện luận văn.

Tôi xin bày tỏ sự biết ơn sâu sắc đến gia đình, người thân, bạn bè đồng nghiệp, những người luôn động viên, khuyến khích và giúp đỡ về mọi mặt để tôi có thể hoàn thành công việc nghiên cứu.

Thái Nguyên, tháng 1 năm 2014

Tác giả luận văn

Vũ Đức Thảo

LỜI CAM ĐOAN

Tôi xin cam đoan:

Những nội dung trong luận văn này là do tôi thực hiện dưới sự hướng dẫn trực tiếp của thầy giáo hướng dẫn TS. Hoàng Đỗ Thanh Tùng

Mọi tham khảo dùng trong luận văn đều được trích dẫn rõ ràng tác giả, tên công trình, thời gian, địa điểm công bố.

Mọi sao chép không hợp lệ, vi phạm quy chế đào tạo, hay gian lận tôi xin chịu hoàn toàn trách nhiệm.

Thái Nguyên, tháng 1 năm 2014
Tác giả luận văn

Vũ Đức Thảo

THUẬT NGỮ VIẾT TẮT

Từ viết tắt	Nghĩa tiếng Anh	Nghĩa tiếng Việt
MAC	Media Access Control	
AP	Access Point	Điểm truy cập
LLC	Logical Link Control	
DW	Data Warehouse	Kho dữ liệu
D		
GPS	Global Position System	Hệ thống định vị toàn cầu
ETL	Extract Transform Load	
IEEE	Institute of Electrical and Electronics Engineers	Viện kỹ nghệ Điện và Điện tử
MAC	Medium Access Layer	Lớp truy nhập môi trường
RF	Radio Frequency	Tần số vô tuyến
RFID	Radio Frequency Identification	Xác định tần số vô tuyến
RTDW	Real time Data Warehouse	Kho dữ liệu thời gian thực
CTF	Capture, Transform and Flow Correctness	
WLAN	Wireless Local Area Network	Mạng cục bộ không dây

MỤC LỤC

MỞ ĐẦU	1
CHƯƠNG 1. TỔNG QUAN VỀ HỆ THỐNG ĐỊNH VỊ TRONG NHÀ... 3	3
1.1 Giới thiệu.....	3
1.2. Một số công nghệ định vị.....	4
1.2.1 Công nghệ hồng ngoại	4
1.2.2 Công nghệ mạng cục bộ không dây (WLAN)	5
1.2.4 Công nghệ quang.....	12
1.3. Một số ứng dụng của hệ thống định vị hiện nay	12
1.3.1 Google maps.....	12
1.3.2 Bing maps.....	14
1.3.3 Nokia Indoor Navigation	15
1.3.4 Fastmall	16
1.3.5 Micello	17
1.3.6 Junaio	18
1.3.7 IndoorAlas.....	19
1.4. Tính thiết yếu của kho dữ liệu thời gian thực cho dữ liệu định vị trong nhà	20
1.5 Mục tiêu của đề tài	23
CHƯƠNG 2. NGHIÊN CỨU VỀ KHO DỮ LIỆU THỜI GIAN THỰC VÀ CÁC VẤN ĐỀ LIÊN QUAN HỆ THỐNG ĐỊNH VỊ, TRỢ GIÚP TRONG NHÀ	24
2.1 Định nghĩa kho dữ liệu.....	24
2.2. Các thành phần kho dữ liệu.....	25
2.2.1. Siêu dữ liệu (Metadata).....	25
2.2.2. Các nguồn dữ liệu	26
2.3. Cơ sở dữ liệu của kho dữ liệu	27
2.3.1 Đặc điểm dữ liệu trong kho dữ liệu	28
2.3.2 Làm sạch và chuyển đổi dữ liệu.....	29
2.3.3 Tích hợp và hợp nhất dữ liệu (ETL)	30

2.3.4 Tải dữ liệu	31
2.4. Các phương pháp lưu trữ dữ liệu (MOLAP, ROLAP, HOLAP)	33
2.4.1. MOLAP (Multidimensional OLAP)	33
2.4.2. ROLAP (Relational OLAP)	33
2.4.3 HOLAP (Hybrid OLAP)	34
2.4.4 Sử dụng kho dữ liệu	35
2.5 Kho dữ liệu thời gian thực	36
2.5.1 Chuẩn hóa và phi chuẩn hóa	36
2.5.2 Dòng Dữ Liệu	37
2.6 Các vấn đề kho dữ liệu định vị trong nhà	39
2.6.1 Các điểm truy cập không dây	39
2.6.2 ETL không dây	39
2.6.3 Vấn đề trùng lặp dữ liệu	40
CHƯƠNG 3. ĐỀ XUẤT MÔ HÌNH TRIỂN KHAI KHO DỮ LIỆU	
THỜI GIAN THỰC CHO HỆ THỐNG ĐỊNH VỊ TRONG NHÀ	41
3.1 Cấu trúc hệ thống	41
3.2 Thiết kế kho dữ liệu thời gian thực	43
3.2.1 Chiều dữ liệu thiết bị kết nối (Bluetooth , Wlan)	44
3.2.2 Chiều dữ liệu thời gian	44
3.2.3 Chiều dữ liệu đối tượng chuyển động	45
3.2.4 Đề xuất thiết kế RTDW cụ thể :	46
3.2.5 Đề xuất thiết kế kho dữ liệu thời gian thực tổng quát	48
3.3 Đề xuất Real time ETL	51
3.3.1 Đề xuất Real time ETL trạng thái	51
3.3.2 Thu nhận dữ liệu thay đổi (Change Data Capture)	53
3.3.3 Chuyển đổi dữ liệu	55
3.3.4 Dò tìm và loại bỏ trả về thời gian thực	55
3.3.5 Lưu lượng	56
KẾT LUẬN	57
TÀI LIỆU THAM KHẢO	58

DANH MỤC HÌNH

Hình 1.1: Các thành phần cơ sở hạ tầng LBS	3
Hình 1.2.3-1 Các ngăn của Bluetooth.	10
Hình 1.2.3-2 Một ví dụ Piconet của Bluetooth gồm một thiết bị chủ (Master) và bốn thiết bị con (Slave).	11
Hình 1.3.1 : Các tính năng trong Indoor Google Maps	12
Hình 1.3.2: Ứng dụng Bing maps	14
Hình 1.3.3: Hình ảnh về một số tính năng nổi bật của Nokia.....	15
Hình 1.3.4: Hình ảnh trong ứng dụng Fastmall	16
Hình 1.3.5: Hình ảnh tính năng dẫn đường của Micello.....	17
Hình 1.3.6: Tính năng nổi bật trong Junaio sử dụng công nghệ Augmented Reality	18
Hình 1.3.7: Ứng dụng điều hướng trong nhà sử dụng từ trường trái đất.....	19
Hình 2.2. Mô hình kho dữ liệu	25
Hình 2.5.2 Dòng dữ liệu RTDW và DW	38
Hình 3.1 : Kiến trúc Hệ thống.....	42
Hình 3.2.1: Chiều dữ liệu Thiết bị kết nối	44
Hình 3.2.2 : Các phạm vi theo Thời gian trong Ngày và Thứ	45
Hình 3.2.3:Chiều dữ liệu đối tượng chuyển động.....	46
Hình 3.2.4: Đề xuất Thiết kế RTDW cụ thể	47
Hình 3.2.5 : Đề xuất thiết kế RTDW chung.....	49
Hình 3.3.1 : Thiết kế mức độ cao của Real time ETL	52
Hình 3.3.2 : Giảm đồ của bảng thực tế	53

MỞ ĐẦU

1. Đặt vấn đề

Trong nền kinh tế hiện nay, thông tin là yếu tố sống còn đối với bất kỳ doanh nghiệp nào. Việc nắm bắt thông tin giúp cho các doanh nghiệp có thể hoạch định các chiến lược kinh doanh cho mình một cách chính xác.

Trải qua quá trình hoạt động kinh doanh, các dữ liệu của doanh nghiệp phát sinh ngày càng nhiều nên việc lưu trữ, quản lí và sử dụng kho dữ liệu được đặc biệt quan tâm. Cùng với sự phát triển mạnh mẽ của Công nghệ thông tin, công nghệ kho dữ liệu (DW-Data Warehouse) ra đời, đáp ứng được nhu cầu quản lý, lưu trữ thông tin có khối lượng lớn và có khả năng khai thác dữ liệu đa chiều và theo chiều sâu nhằm hỗ trợ việc ra quyết định của các nhà quản lý. Với các DW thông thường việc nạp dữ liệu trong khoảng thời gian cố định, thường là một lần mỗi ngày, tuần, hoặc tháng sẽ mất nhiều thời gian và không đồng bộ kịp dữ liệu. Vì vậy cần thiết phải có một hệ thống DW không chỉ đơn giản là lưu trữ dữ liệu, đưa ra kết quả chính xác mà nó cần phải đảm bảo thực hiện một xử lý trong một thời gian rất ngắn, người ta gọi đó là hệ thống kho dữ liệu (thông tin) thời gian thực.

Hệ thống thông tin thời gian thực ngày nay được ứng dụng trong rất nhiều lĩnh vực như giao dịch cổ phiếu chứng khoán , đặt chỗ máy bay , du lịch , tính cước trả trước viễn thông... Với sự phát triển nhanh chóng của mạng Internet, viễn thông và các dịch vụ toàn cầu hóa, đòi hỏi phải có những hệ thống kho dữ liệu (thông tin) thời gian thực để đáp ứng được các yêu cầu của người dùng.

Trong phạm vi nghiên cứu của luận văn học viên nghiên cứu về:

“GIẢI PHÁP KHO DỮ LIỆU THỜI GIAN THỰC CHO HỆ THỐNG SIÊU THỊ ”

Cụ thể trong nội dung luận văn học viên sẽ nghiên cứu về kho dữ liệu thời gian thực cho dữ liệu định vị trong nhà nhằm giúp người quản lý nắm bắt được sự quan tâm của khách hàng đối với các gian hàng cũng như điều chỉnh được mật độ lưu thông của khách hàng trong siêu thị tránh tắc nghẽn

2. Mục tiêu nghiên cứu của đề tài

Đề xuất một thiết kế kho dữ liệu thời gian thực cho hệ thống định vị trong nhà

3. Đối tượng và phạm vi nghiên cứu

Nghiên cứu thiết kế kho dữ liệu thời gian thực cho hệ thống định vị trong nhà có thể áp dụng trong các trung tâm thương mại lớn hoặc tại các sân bay quản lý thông tin dữ liệu di chuyển của khách hàng

4. Hướng nghiên cứu của đề tài

Nghiên cứu về phương pháp xây dựng và cấu trúc kho dữ liệu thời gian thực

Nghiên cứu về công nghệ Bluetooth , Wlan, wifi

Nghiên cứu các bài toán về ETL, Bounce ...

5. Những nội dung nghiên cứu chính

Luận văn được trình bày 3 chương, các nội dung cơ bản của luận văn được trình bày theo cấu trúc như sau:

Mở đầu

Chương 1 : Cơ sở lý thuyết

Tổng quan về Hệ thống định vị trong nhà

Chương 2 : Nghiên cứu về kho dữ liệu thời gian thực và các vấn đề liên quan cho hệ thống định vị, trợ giúp trong nhà

Chương 3 : Đề xuất mô hình triển khai kho dữ liệu thời gian thực cho hệ thống định vị trong nhà

Kết luận và hướng phát triển của luận văn