

ỨNG DỤNG KHO DỮ LIỆU VÀ TRỰC QUAN HÓA DỮ LIỆU TRONG QUẢN LÝ SỐ LIỆU NUÔI TRỒNG THỦY SẢN TẠI TRÀ VINH

APPLIED DATA WAREHOUSE AND DATA VISUALIZATION IN AQUACULTURE DATA MANAGEMENT IN TRA VINH PROVINCE

Nguyễn Bảo Ân¹, Nguyễn Nhứt Lam², Hà Thị Thúy Vi³

Tóm tắt – Bài báo mô tả quá trình xây dựng một trang web quản lý báo cáo số liệu nuôi trồng thủy sản trên địa bàn tỉnh Trà Vinh. Trên đó, một kho dữ liệu được xây dựng để hỗ trợ cho việc phân tích và trực quan hóa số liệu nhằm đưa ra các báo cáo và biểu đồ hỗ trợ cho việc nắm bắt tình hình nuôi trồng thủy sản và ra quyết định của các cấp quản lý. Hệ thống được xây dựng với công nghệ .NET, hệ quản cơ sở dữ liệu Microsoft SQL Server 2014, các kỹ thuật phân tích dữ liệu OLAP trong Microsoft SQL Analysis Services 2014, Reporting services 2017 và tiện ích OLAP của Microsoft Excel 2013.

Từ khóa: kho dữ liệu, OLAP, trực quan hóa, thủy sản Trà Vinh

Abstract – This paper introduces an information system developed to maintain data management and reporting on aquaculture in Tra Vinh province. A data warehouse was also built for data analysis and visualization in order to produce general reports and charts that help the managers in decision making. The proposed system was built using .NET technology, Microsoft SQL Server 2014 DBMS, Microsoft SQL Analysis Services and OLAP features of Microsoft Excel 2013.

Keywords: data warehouse, OLAP, data visualization, Tra Vinh aquaculture.

^{1,2,3}Bộ môn Công nghệ Thông tin, Khoa Kỹ thuật và Công nghệ, Trường Đại học Trà Vinh
Email: annb@tvu.edu.vn
Ngày nhận bài: 26/10/2017; Ngày nhận kết quả bình duyệt: 05/12/2017; Ngày chấp nhận đăng: 29/12/2017

I. MỞ ĐẦU

Trà Vinh là một tỉnh ven biển thuộc Đồng bằng sông Cửu Long. Trà Vinh có lợi thế lớn về nông nghiệp nói chung và nuôi trồng thủy sản nói riêng. Hiện nay, các loại thủy sản được nuôi trồng trên địa bàn tỉnh rất đa dạng gồm: tôm sú, tôm thẻ chân trắng, tôm càng xanh, cua biển, cá lóc, cá tra, cá nước ngọt các loại, nghêu,... Hình thức nuôi cũng đa dạng bao gồm nuôi công nghiệp, thâm canh, bán thâm canh, quảng canh; tôm sú và tôm càng xanh còn có hình thức nuôi kết hợp như tôm rừng (nuôi trong rừng ngập mặn), tôm lúa (nuôi tôm kết hợp với lúa), tôm cá (nuôi tôm kết hợp với cá), tôm cua (nuôi tôm kết hợp với nuôi cua)... Tuy nhiên, cùng với xu hướng tăng trưởng quy mô thả nuôi, thiệt hại trong nuôi thủy sản trong các năm gần đây cũng tăng dần, đặt ra nhu cầu về chiến lược quản lý đối với lãnh đạo ngành thủy sản nói riêng và ngành nông nghiệp nói chung.

Nhiệm vụ quản lý số liệu thủy sản hiện nay được giao cho Chi cục Thủy sản trực thuộc Sở Nông nghiệp và Phát triển Nông thôn tỉnh. Việc báo cáo số liệu được thực hiện theo chu kỳ hằng tuần, hằng tháng, hằng quý và hằng năm theo cơ chế phân cấp: cấp xã báo cáo lên cấp huyện, cấp huyện báo cáo lên cấp tỉnh. Công tác tổng hợp báo cáo ở cấp xã được thực hiện bởi cán bộ nông nghiệp xã, ở cấp huyện được thực hiện bởi cán bộ Phòng Nông nghiệp huyện, và ở cấp tỉnh được thực hiện bởi cán bộ kỹ thuật của Chi cục Thủy sản.

Với quy trình hiện nay, do các cán bộ phụ trách báo cáo số liệu chưa được trang bị phần mềm chuyên dụng nên tất cả các số liệu đều được nhập và tính toán bằng phần mềm Microsoft Excel, sau

đó, cán bộ mới lập một báo cáo văn bản dựa vào số liệu trong tập tin excel để báo cáo lên cấp trên qua e-mail. Sau khi nhận được e-mail báo cáo, cán bộ phụ trách số liệu ở cấp trên tính toán tổng hợp số liệu lại từ tất cả các địa phương trực thuộc bằng excel và tiếp tục soạn thảo văn bản báo cáo dựa vào số liệu tính toán được. Một số hạn chế của quy trình hiện hành có thể liệt kê sau đây: i) gây mất thời gian, có nguy cơ chậm trễ trong việc gửi báo cáo cho Ủy ban Nhân dân tỉnh hoặc Tổng cục Thủy sản; ii) có thể gây sai sót về số liệu khi thống kê, do số liệu của các địa phương được lưu trữ trong những tập tin excel rời rạc, nên tổng hợp thủ công số liệu từ nhiều tập tin là một việc làm nhàm chán, dễ gây sai sót và không dễ thực hiện; iii) các bản lưu số liệu không giúp trả lời được các câu hỏi về so sánh, xu hướng,... (ví dụ: trong ba tháng vừa qua, diện tích nuôi trồng thủy sản ở huyện Cầu Ngang tăng hay giảm, so với cùng kì năm trước như thế nào?); iv) để nắm được tình hình chung về diện tích thả nuôi, số lượng con giống, số lượng thu hoạch, số lượng thiệt hại, tình hình dịch bệnh,... cán bộ quản lí ở cấp huyện, cấp tỉnh phải đọc hết tất cả báo cáo văn bản và tự tổng hợp các thông tin này hoặc tính toán từ hàng chục tập tin excel rời rạc chứ không có một phương tiện trực quan như biểu đồ, dashboard,... để trợ giúp ra quyết định.

Trước những vấn đề trên, nhằm nâng cao hiệu quả và chất lượng công tác báo cáo số liệu nuôi trồng thủy sản ở tỉnh nhà, chúng tôi xây dựng hệ thống báo cáo số liệu trực tuyến kết hợp công nghệ data warehouse (kho dữ liệu) và trực quan hóa dữ liệu cho công tác báo cáo số liệu trên địa bàn tỉnh Trà Vinh.

II. TỔNG QUAN NGHIÊN CỨU

A. Kho dữ liệu

1) *Định nghĩa:* Từ những năm 1970, Bill Inmon, cha đẻ của thuật ngữ kho dữ liệu (data warehouse) đã định nghĩa kho dữ liệu là một tập hợp dữ liệu tương đối ổn định (không hay thay đổi), cập nhật theo thời gian, tích hợp theo hướng chủ đề được sử dụng chủ yếu trong việc hỗ trợ quá trình ra quyết định về mặt quản lí [1]. Về mặt vận hành, kho dữ liệu thường được duy trì tách biệt với cơ sở dữ liệu của hệ thống. Trong khi cơ sở dữ liệu hệ thống thường hỗ trợ xử lí các giao

dịch trực tuyến (OnLine Transaction Processing – OLTP), kho dữ liệu lại hỗ trợ cho thao tác phân tích dữ liệu trực tuyến (OnLine Analytical Processing – OLAP). Các cơ sở dữ liệu dành cho OLTP thường hỗ trợ các tác vụ xử lí dữ liệu thường xuyên xảy ra trong quá trình hoạt động của hệ thống nên liên tục có các thao tác đọc, ghi và cập nhật dữ liệu thường xuyên, chẳng hạn các giao dịch của một ngân hàng, một hệ thống thương mại điện tử,... Vì thế, cơ sở dữ liệu OLTP thường phải được chuẩn hóa (normalized) để tối ưu hiệu năng truy vấn và cập nhật, trong đó phải luôn bảo đảm tính nhất quán và khả năng khôi phục dữ liệu. Trái lại, do được xây dựng nhằm mục tiêu phân tích dữ liệu lịch sử để hỗ trợ ra quyết định [2] nên kho dữ liệu không tập trung vào tính cập nhật tức thời của dữ liệu mà tối ưu hóa thao tác truy vấn cần tính toán dựa trên số lượng rất lớn các bản ghi, để trả lời các truy vấn mang tính tổng hợp, ví dụ: tổng doanh thu trong 5 năm gần đây cao hay thấp hơn so với 5 năm trước đó... Chi tiết so sánh giữa OLTP và OLAP được thể hiện trong Bảng 1.

2) *Kiến trúc kho dữ liệu:* Kiến trúc của kho dữ liệu, xét theo tính năng, có thể chia thành ba tầng: tầng đáy – data warehouse server, tầng trung gian – OLAP server, và tầng đỉnh – các công cụ đầu cuối (front-end tools) [3]. Khi có một lượng lớn dữ liệu cần đưa vào kho, người ta dùng các công cụ ETL (Extract – Transform – Load) để trích chọn dữ liệu từ nhiều nguồn, trích chọn và chuyển định dạng các dữ liệu này về một định dạng thống nhất và nạp vào kho. Về mặt vật lí, kho dữ liệu là một cơ sở dữ liệu được thiết kế đặc biệt (theo lược đồ hình sao hoặc bông tuyết) để đáp ứng nhu cầu tính toán và truy vấn với số lượng lớn. Một kho dữ liệu có thể bao gồm nhiều data mart, tức những tập hợp dữ liệu có giá trị với một nhóm người dùng nhất định. Dữ liệu trong kho sẽ được xử lí qua một công cụ OLAP để kết xuất kết quả ra cho các công cụ đầu cuối như truy vấn, báo cáo, phân tích và khai thác dữ liệu. Mô hình kiến trúc của kho dữ liệu được thể hiện trong Hình 1.

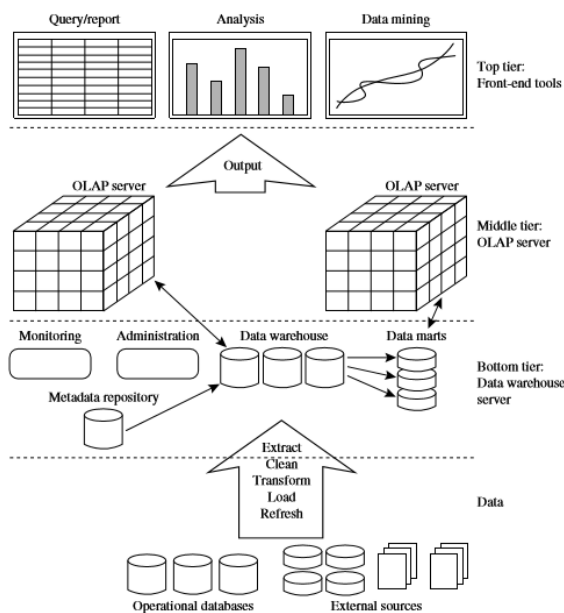
B. Data cube và OLAP

Một kho dữ liệu được hình thành dựa trên một mô hình dữ liệu đa chiều, mô hình này thể hiện

Bảng 1. So sánh giữa OLTP và OLAP

Tiêu chí	OLTP (Online Transaction Processing)	OLAP (Online Analysis Processing):
Người dùng	Nhân viên, bộ phận IT, khách hàng hàng ngàn người	Chuyên viên phân tích dữ liệu, bộ phận ra quyết định vài chục đến vài trăm người
Dữ liệu nguồn	Được nhập liệu thường xuyên bởi những người dùng cuối. Dữ liệu chi tiết tới từng nghiệp vụ	Dữ liệu hợp nhất. Thường được lưu trữ dưới dạng đã được tính toán, có tính lịch sử và theo mô hình đa chiều (multi-dimensional)
Mục đích của dữ liệu	Để kiểm soát và chạy các nghiệp vụ kinh tế cơ bản, thiết yếu xảy ra hàng ngày. Phục vụ người dùng ở cấp hoạt động trong doanh nghiệp	Phục vụ mục đích lập kế hoạch, giải quyết vấn đề, hỗ trợ ra quyết định. Phục vụ cho các cấp bậc quản lý chiến lược.
Phản ánh dữ liệu nào	Cho thấy một ảnh chụp nhanh của các quá trình kinh doanh liên tục	Số liệu cung cấp góc nhìn đa chiều các hoạt động kinh doanh trong doanh nghiệp.
Thêm và sửa dữ liệu	Thêm, sửa nhanh chóng và được tạo bởi người dùng cuối ở cấp hoạt động, một số lượng lớn giao dịch (insert, update, delete) trong một thời gian ngắn	Chạy theo chu kỳ, theo lô lớn các giao dịch, như vào một giờ nhất định để tập hợp, tính toán và làm mới dữ liệu. Số lượng tương đối thấp của các giao dịch. Thường có nhiều truy vấn yêu cầu tính toán tổng hợp dữ liệu.
Truy vấn	Liên quan đến các truy vấn cơ bản, số lượng dòng dữ liệu trả về ít.	Thường truy vấn với các câu truy vấn tính toán phức tạp dựa trên khối lượng dữ liệu rất lớn.
Tốc độ xử lý	Thường rất nhanh trong một vài giây	Phụ thuộc vào khối lượng dữ liệu cần xử lý. Một lần xử lý một lô query để tập hợp và tính toán dữ liệu có thể kéo dài hàng giờ.
Không gian lưu trữ	Tương đối nhỏ nếu dữ liệu lịch sử được cắt giảm và lưu trữ thích hợp (trăm MB - GB)	Thường rất lớn, do sự tồn tại của dữ liệu tổng hợp mang tính lịch sử, cần nhiều chỉ mục hơn so với OLTP (trăm GB - TB)
Thiết kế database	Mô hình thiết kế theo mô hình quan hệ thực thể, cơ sở dữ liệu thiết kế cho hệ thống này thường được chuẩn hóa	Thường không cần chuẩn hóa, ít table hơn. Thường theo lược đồ hình sao hoặc bông tuyết
Sao lưu và phục hồi dữ liệu	Dữ liệu hoạt động là quan trọng với hoạt động nên phải bảo trì sao lưu cẩn thận, việc mất dữ liệu đồng nghĩa với việc mất thời gian, tiền bạc và công sức nhập liệu, trách nhiệm pháp lý.	Khi mất mát dữ liệu xảy ra, nhiều hệ thống OLAP thường load lại dữ liệu từ các nguồn dữ liệu gốc ở các hệ thống OLTP như là một giải pháp phục hồi dữ liệu.

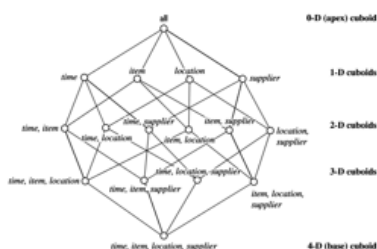
(Nguồn: *Data Mining: Concepts and Techniques [2]*)



Hình 1: Kiến trúc 3 tầng của kho dữ liệu
(Nguồn: *An Overview of Data Warehousing and OLAP Technology*)

dữ liệu dưới dạng những data cube (khối dữ liệu). Một data cube cho phép mô hình hóa và xem dữ liệu trên nhiều chiều, vì thế mô hình dữ liệu đa chiều có hai loại bảng: i) bảng chiều (dimension table): thể hiện thông tin về một chiều nào đó của dữ liệu, ví dụ: thời gian (ngày, tháng, năm), địa điểm (thành phố, vùng, quốc gia),...; ii) bảng fact (fact table) chứa các thuộc tính quan tâm (measures) của cube như số lượng bán ra, tổng thành tiền bán ra,... và khóa ngoại liên kết với các bảng chiều. Hình 2a là ví dụ về sự kết hợp giữa các chiều trong một kho dữ liệu của một hệ thống thương mại điện tử có bốn chiều: thời gian (time), sản phẩm (item), vị trí (location) và nhà cung cấp (supplier). Hình 2b thể hiện một data cube với ba chiều thời gian, sản phẩm và vị trí, và measure là số lượng bán ra.

Trong một cơ sở dữ liệu OLAP, các thao tác được sử dụng nhiều nhất là: roll-up (cuộn lên), drill-down (khoan xuống), slice and dice (cắt lát) và pivot (quay). Trong đó, roll-up là thao tác loại bỏ một chiều nào đó ra khỏi cube, và số liệu



a) Sự kết hợp giữa các chiều trong data cube

	Chicago	854	882	89	623
New York	1087	968	38	872	
Toronto	818	746	43	591	
Vancouver					
Q1	605	825	14	400	682, 925, 789
Q2	680	952	31	512	728, 1002, 789
Q3	812	1023	30	501	798, 984, 870
Q4	927	1038	38	580	
	computer	phone	security	home entertainment	
	item (types)				

b) Data cube 3 chiều

Hình 2: Cấu trúc của data cube

(Nguồn: *Data Mining: Concepts and Techniques* [2])

theo các chiều còn lại sẽ được tính toán lại theo một hàm tóm lược (aggregation function) nào đó. Drill-down là thao tác ngược lại của roll-up, khi ta muốn chi tiết hóa số liệu theo một chiều nào đó từ số liệu đã được tóm lược. Slice là thao tác trích chọn dữ liệu theo một chiều nào đó, còn dice là thao tác trích chọn dữ liệu theo hai chiều hoặc hơn. Pivot là thao tác trình bày dữ liệu trực quan khi trình bày dữ liệu trong một bảng chữ nhật, trong đó, các cạnh là các chiều của dữ liệu.

C. Tình hình nghiên cứu

Việc ứng dụng công nghệ kho dữ liệu vào hệ thống thông tin phục vụ nông nghiệp đã được nghiên cứu rộng rãi trên thế giới. Các mô hình hệ thống thường được đề xuất nhằm quản lý số liệu về tình hình sản xuất, thu hoạch và dịch bệnh,... Năm 2000, Yost [4] đề xuất mô hình về kho dữ liệu và hệ thống hỗ trợ quyết định cho sử dụng số liệu thống kê cho Bộ Nông nghiệp Hoa Kỳ. Gupta và cộng sự [5] đề xuất mô hình cơ sở dữ liệu đa

chiều cho kho dữ liệu trong nông nghiệp,... Hầu hết các nghiên cứu về xây dựng kho dữ liệu cho lĩnh vực nông nghiệp đều nhằm đưa ra một mô hình hệ thống và mô hình cơ sở dữ liệu để tích hợp các dữ liệu nông nghiệp trong quá khứ vào một kho dữ liệu, sau đó sử dụng các công cụ phân tích dữ liệu trực tuyến (OLAP) hoặc khai khoáng dữ liệu (data mining) để rút trích ra các mẫu, các tri thức tiềm ẩn bên dưới dữ liệu mà người sở hữu dữ liệu khó có thể đúc kết, mừng tượng ra được nếu thiếu công cụ, từ đó sử dụng các tri thức này cho việc hỗ trợ quyết định.

Đối với trong nước, hệ thống báo cáo số liệu trực tuyến không phải là mới đối với các cơ quan nhà nước. Hệ thống này đã được nhiều cơ quan, ban ngành triển khai. Tuy nhiên, thao tác tích hợp dữ liệu vào data warehouse và trực quan hóa dữ liệu, hỗ trợ lãnh đạo ra quyết định cho nghiệp vụ chưa được quan tâm đúng mức. Trong lĩnh vực nông nghiệp, Trường Đại học Cần Thơ đã thực hiện một đề tài có phạm vi nghiên cứu rất rộng về hệ thống thông tin hỗ trợ phòng chống dịch bệnh cây trồng và thủy sản cho vùng kinh tế trọng điểm [6]. Đề tài đưa ra rất nhiều giải pháp về cổng thông tin, hệ thống báo cáo, phân tích dữ liệu trực tuyến (OLAP), web ngữ nghĩa để dự đoán, chẩn đoán bệnh lúa và bệnh thủy sản, hệ thống thông tin địa lý (GIS), mô phỏng,... Mục đích của đề tài nhằm đưa ra một giải pháp tổng thể về các hệ thống thông tin xử lý dịch bệnh trong nông nghiệp chứ không nhằm phục vụ cho công tác quản lý số liệu của các cơ quan nhà nước. Năm 2012, Phan Quốc Nghĩa [7] đề xuất xây dựng một mô hình cơ sở dữ liệu đa chiều và phân tích OLAP trên tình hình dịch bệnh trên lúa ở Trà Vinh. Năm 2013, Sở Nông nghiệp và Phát triển Nông thôn Thành phố Cần Thơ [8] đã phát triển hệ thống WebGIS quản lý dữ liệu cho sáu chi cục trực thuộc (Thủy sản, Thủy lợi, Thú y, Phát triển Nông thôn, Bảo vệ Thực vật, Nước sạch và Vệ sinh Môi trường) để tận dụng chức năng của hệ thống thông tin địa lý vào việc quản lý số liệu nông nghiệp. Năm 2013, Phạm Thị Xuân Lộc và cộng sự [9] nghiên cứu xây dựng kho dữ liệu lúa-tôm-cá cho khu vực Đồng bằng sông Cửu Long ứng dụng các kỹ thuật datamart và OLAP. Năm 2016, Bộ Nông nghiệp và Phát triển Nông thôn [10], được sự hỗ trợ từ chính phủ

Phần Lan, đã triển khai dự án Phát triển hệ thống thông tin quản lý ngành Lâm nghiệp tại Việt Nam, với các chức năng quản lý số liệu lâm nghiệp và số liệu kinh tế xã hội của các vùng dân cư có rừng, hỗ trợ phân tích và ra quyết định quản lý trong lĩnh vực lâm nghiệp. Các nghiên cứu trên thể hiện việc phát triển các hệ thống thông tin quản lý nông nghiệp là nhu cầu rất thiết thực. Vì vậy, với lĩnh vực thủy sản, việc phát triển một hệ thống báo cáo số liệu tích hợp các tính năng của kho dữ liệu và kinh doanh thông minh (Business Intelligence) sẽ có tính khả thi và ứng dụng cao.

III. NỘI DUNG

A. Mô hình hệ thống

Với nhu cầu phát triển một hệ thống báo cáo có ứng dụng các công cụ phân tích và trực quan hóa số liệu nuôi trồng thủy sản như đã mô tả ở phần I, chúng tôi đề xuất một mô hình hệ thống như Hình 3. Hệ thống bao gồm một website dành cho cán bộ nông nghiệp cấp xã, huyện và cán bộ Chi cục Thủy sản báo cáo số liệu hàng tuần; trong đó, quan trọng nhất là báo cáo tuần của cấp xã. Báo cáo bao gồm các số liệu thay đổi trong tuần về tình hình nuôi trồng thủy sản trên địa bàn của mình phụ trách, để làm cơ sở cho các báo cáo của cán bộ cấp cao hơn. Cơ sở dữ liệu của hệ thống báo cáo được đặt tên là QLTS_DB. Hàng tuần, dữ liệu báo cáo cũng được nạp vào kho dữ liệu (QLTS_DW) để làm cơ sở cho các thao tác phân tích và báo cáo trực quan. Cơ sở dữ liệu của website và kho dữ liệu được phát triển trên hệ quản trị cơ sở dữ liệu MS SQL Server 2014, công cụ OLAP được hỗ trợ bởi MS SQL Analysis Services 2014, các công cụ front-end là MS Reporting Services 2017 và Microsoft Excel 2013. Chúng tôi cũng phát triển một module ETL để nạp dữ liệu từ các tập tin excel chứa báo cáo từ vụ mùa 2014-2015 đến nay và từ cơ sở dữ liệu báo cáo vào kho dữ liệu. Chi tiết phát triển các thành phần hệ thống được mô tả trong các mục B và C của phần này.

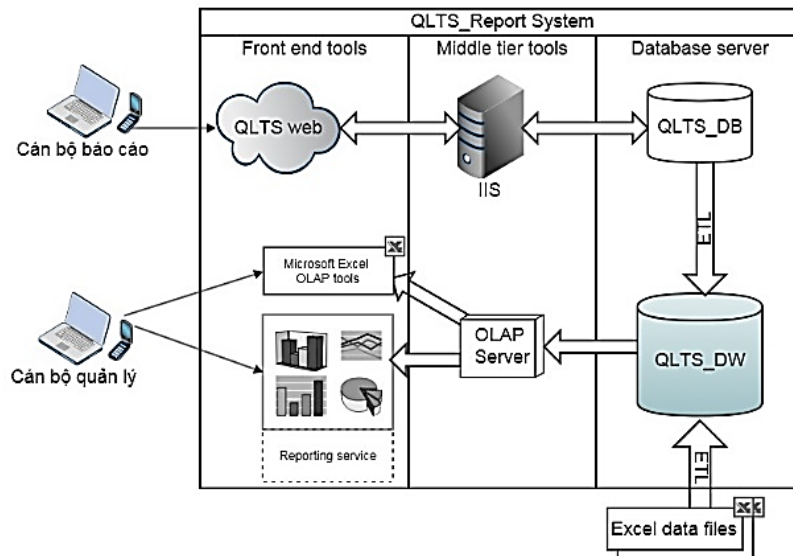
B. Phát triển hệ thống báo cáo số liệu trực tuyến

Dựa vào cơ cấu nuôi trồng thủy sản và đơn vị hành chính của tỉnh, cơ sở dữ liệu cho hệ thống báo cáo số liệu theo tuần, tháng, năm được thiết kế như Hình 4. Để tiện lợi trong việc nạp dữ

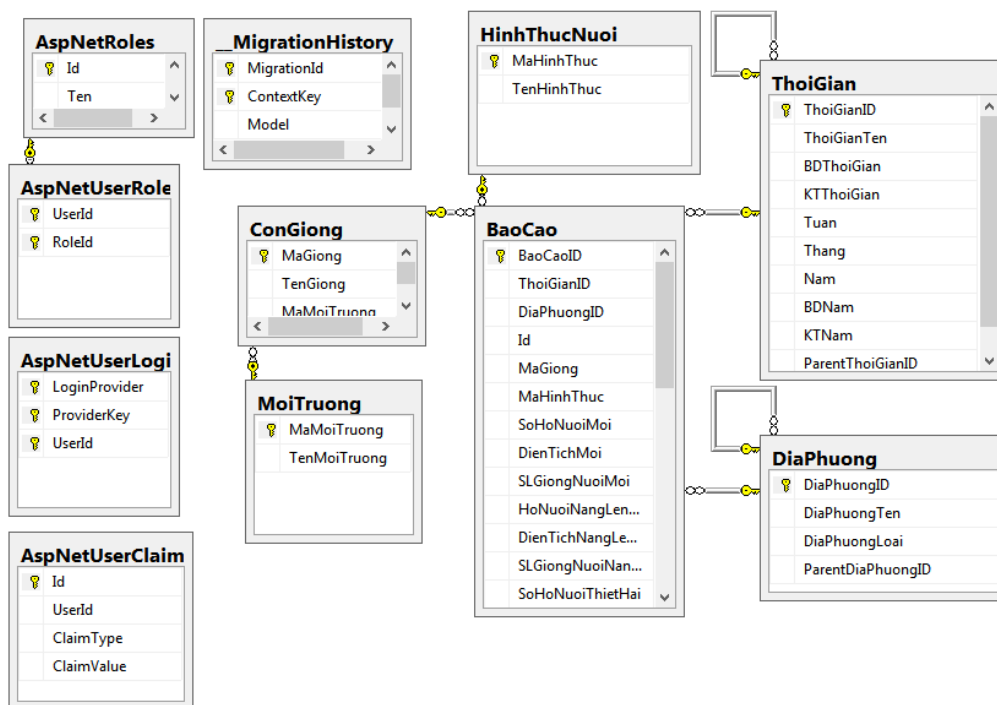
liệu từ cơ sở dữ liệu báo cáo (QLTS_DB) nạp sang kho dữ liệu (QLTS_DW), cơ sở dữ liệu của website báo cáo cũng được thiết kế theo lược đồ hình sao như một kho dữ liệu. Trong đó, số liệu phát sinh trong một tuần về một loại thủy sản của một địa phương sẽ được lưu trong một bản ghi của bảng BaoCao (báo cáo), các số liệu cần nhập bao gồm: sản lượng thu hoạch (tấn), số hộ, diện tích (hecta), số lượng giống (ngàn con) được thả nuôi và bị thiệt hại trong tuần; sau đó mỗi số liệu chi tiết sẽ được tính toán số nâng lên (tổng số tích lũy từ đầu vụ đến tuần hiện hành), và tỉ lệ (%) giữa thiệt hại so với thả nuôi. Các loại thủy sản (bảng ConGiong) cũng được phân loại theo môi trường thả nuôi (bảng MoiTruong) bao gồm môi trường nước ngọt hoặc nước mặn-lợ. Hình thức thả nuôi (bảng HìnhThucNuoi, bao gồm công nghiệp, bán công nghiệp và quảng canh) cũng được bao gồm trong báo cáo. Tính phân cấp của báo cáo theo đơn vị hành chính thể hiện trong bảng địa phương (DiaPhuong). Trong đó, địa phương được phân theo ba cấp: xã, huyện và tỉnh. Chiều thời gian của báo cáo được thể hiện trong bảng ThoiGian và phân cấp theo ngày, tuần, tháng, năm và mùa vụ. Website báo cáo được phát triển sử dụng công nghệ ASP.NET MVC nên cơ sở dữ liệu cũng được tự động chèn thêm một số bảng khác phục vụ cho thao tác phân quyền và xác thực người dùng.

Website báo cáo có ba cấp độ người dùng tương ứng với ba cấp độ quản lý theo đơn vị hành chính (xã, huyện, tỉnh). Mỗi người dùng có quyền tạo và cập nhật báo cáo định kỳ (theo tuần, tháng, năm) thuộc địa bàn mình phụ trách và được quyền xem báo cáo của các địa phương trực thuộc địa bàn mình quản lý. Như vậy, cán bộ cấp tỉnh có thể xem tất cả các báo cáo từ cấp xã, huyện đến cấp tỉnh. Chi tiết về phân quyền của hệ thống được thể hiện trong Bảng 2.

Khi hệ thống vận hành, chỉ duy nhất báo cáo tuần của cấp xã là đòi hỏi cán bộ nông nghiệp xã phải thu thập số liệu và nhập liệu vào hệ thống, các báo cáo còn lại sẽ được tổng hợp tự động từ số liệu sơ cấp của báo cáo tuần cấp xã. Các cấp báo cáo có giao diện tương đối giống nhau. Hình 5 thể hiện giao diện báo cáo tuần của cấp huyện.



Hình 3: Mô hình hệ thống báo cáo số liệu nuôi trồng thủy sản



Hình 4: Mô hình cơ sở dữ liệu website báo cáo số liệu nuôi trồng thủy sản

Bảng 2: Phân quyền truy cập hệ thống báo cáo

Quyền	Báo cáo xã		Báo cáo huyện		Báo cáo tỉnh	
	Tạo/Sửa	Xem	Tạo/Sửa	Xem	Tạo/Sửa	Xem
CB						
Cán bộ xã	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				
Cán bộ huyện		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
Cán bộ tỉnh		<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Báo cáo Tuần 37 tháng 7 năm 2016

Tuần Trước Show 25 entries Search:

Con giống	Thả nuôi						Thiệt hại						Sản lượng				
	Trong tuần			Năng lên			Trong tuần			Năng lên			Trong tuần	Năng lên			
	Hệ nuôi	Diện tích (ha)	Số lượng giống (ngàn)	Hệ nuôi	Diện tích (ha)	Số lượng giống (ngàn)	Hệ nuôi	Diện tích (ha)	Số lượng giống (ngàn)	Hệ nuôi	% so với thả nuôi	Diện tích (ha)	% so với thả nuôi	Số lượng giống (ngàn)	% so với thả nuôi	Sản lượng trong tuần (tấn)	Tổng sản lượng (tấn)
Huyện Cầu Ngang																	
Cá các loại	0	0	0	376	25.3	5427	0	0	0	0	0%	0	0%	0	0%	46	102.7
Cá lóc	2	1.4	353	13	11.05	2445	0	0	0	0	0%	0	0%	0	0%	0	151.6
Cua biển	0	0	0	59	24.6	8135.8	0	0	0	0	0%	0	0%	0	0%	0	0
Nghêu	0	0	0	2	95	64.6	0	0	0	0	0%	0	0%	0	0%	0	0
Tôm chân trắng	173	86.61	44911	4487	1889.93	893263	56	23.19	12298	1277	28.46%	459.14	24.29%	220816	24.72%	0	0
Tôm sú	106	80.22	12616	2748	1398.28	264766	22	10.5	1929	896	32.61%	392.5	28.07%	70435	26.6%	0	0
Cá thác lát																	
Cá tra																	
Tôm càng xanh																	
Tôm phi CN																	

Showing 1 to 10 of 10 entries Previous 1 Next

Hình 5: Giao diện tạo báo cáo tuần cấp huyện

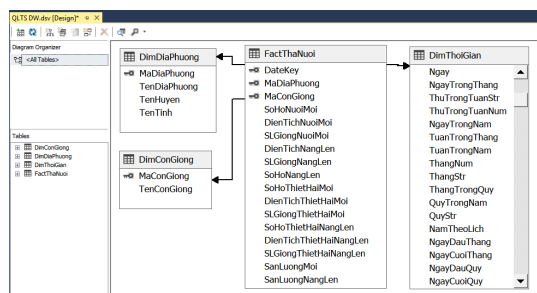
C. Xây dựng kho dữ liệu và công cụ báo cáo

1) *Mô hình cơ sở dữ liệu đa chiều*: Mô hình cơ sở dữ liệu đa chiều thể hiện các góc nhìn data cube cung cấp cho người dùng. Các chiều có thể được thêm/bớt hoặc co/dãn để trả lời các truy vấn từ tổng hợp đến chi tiết. Ví dụ ta có yêu cầu: i) Cho biết số liệu về thả nuôi tôm sú trong toàn tỉnh trong năm 2017 và ii) Cho biết số liệu về thả nuôi tôm sú của các xã thuộc huyện Cầu Ngang trong tuần 22 năm 2017. Trong truy vấn thứ i), ta cần số liệu tổng quát, thông tin về địa phương và thời gian được “co” lại đến mức tổng quát nhất là năm, số liệu trả về là kết quả tổng hợp dựa trên số liệu của tất cả các địa phương trong cả năm. Trong câu truy vấn thứ ii), để thấy chiều thời gian đã được cắt lát tại tháng 8 và chiều địa phương dẫn ra đến cấp xã và được lọc bởi tên huyện, số liệu trình bày được trích chọn từ các báo cáo tuần của cấp xã.

Trên thực tế, số liệu báo cáo cần được tổng hợp/trích chọn trên ba chiều: thời gian, địa phương và con giống. Trong đó, địa phương được phân theo ba cấp hành chính là xã, huyện và tỉnh.

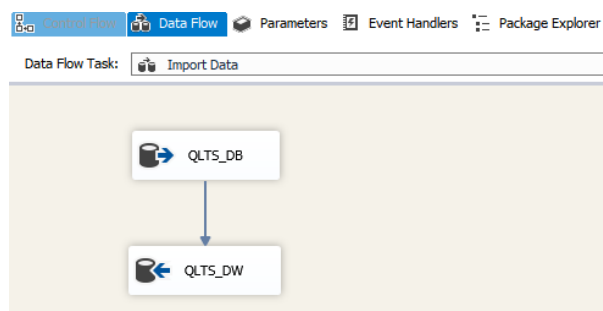
Do đặc thù định kỳ báo cáo nên chiều thời gian được phân cấp nhỏ nhất đến cấp tuần mùa vụ, cao hơn là tháng mùa vụ và năm mùa vụ (mỗi mùa vụ bắt đầu từ tháng 11 của năm trước đến hết tháng 10 của năm sau). Chiều con giống được thiết kế đơn giản nhất, do không có cây phân cấp khái niệm trên chiều này. Mô hình cơ sở dữ liệu được thiết kế theo lược đồ hình sao như trong Hình 6. Trong đó, bảng sự kiện (fact table) được liên kết với các bảng chiều thông qua khóa ngoại và chứa các thông tin về thả nuôi, thiệt hại và sản lượng hàng tuần của các địa phương trên từng loại con giống thả nuôi. Dựa trên cấu trúc đó, các data cubes trong OLAP server (SQL Server Analysis Services) được thiết kế để phục vụ các nhu cầu truy vấn dựa trên các tiêu chí khác nhau về thời gian, địa phương và con giống. Các bộ dữ liệu (datasets) phục vụ cho việc tạo báo cáo và vẽ biểu đồ cũng được tạo nên nhờ các data cubes này.

2) *Công cụ ETL*: Chúng tôi cũng thiết kế một công cụ ETL để nạp dữ liệu từ cơ sở dữ liệu của website báo cáo (QLTS_DB) vào kho dữ liệu



Hình 6: Mô hình cơ sở dữ liệu đa chiều của hệ thống

(QLTS_DW). Để đơn giản, chúng tôi sử dụng bộ công cụ tích hợp dữ liệu SQL Server Integration Services (SSIS). Công cụ này được kích hoạt hàng tuần để đảm bảo số liệu trong kho được cập nhật. Mô hình luồng dữ liệu của công cụ ETL được thể hiện trong Hình 7.



Hình 7: Lược đồ data flow của công cụ ETL

3) *Kết xuất báo cáo và biểu đồ*: Như đã đề cập trong mô hình hệ thống ở Hình 3, cán bộ quản lý có thể trực tiếp truy vấn dữ liệu bằng công cụ OLAP được tích hợp trong Microsoft Excel hoặc xem các báo cáo được thiết kế sẵn bởi chuyên viên xử lý số liệu thông qua cổng dịch vụ báo cáo SQL Reporting Services.

Với Microsoft Excel, người dùng có thể tạo các PivotTable bằng cách kéo thả các trường của các bảng chiều và bảng sự kiện trong kho dữ liệu để kết xuất số liệu theo dạng một bảng hai chiều, trong đó dữ liệu trên các chiều có thể được lọc, roll-up, drill-down một cách dễ dàng. Hình 8 thể hiện một truy vấn về tình hình thả nuôi của từng con giống trên toàn tỉnh với chiều thời gian được lọc trên mùa vụ 2015-2016, chiều địa phương có thể được co giãn đến cấp huyện. Tiện ích này rất

mạnh mẽ và thân thiện, nhờ đó, người dùng có thể dễ dàng quan sát số liệu và lập báo cáo theo từng nhu cầu cụ thể mà không cần phải thực hiện thủ công việc tìm kiếm và tổng hợp số liệu.

Với thao tác tương tự như tạo PivotTable, để quan sát một cách trực quan biến động của tình hình thả nuôi hoặc thiệt hại, so sánh các tiêu chí, người dùng có thể sử dụng công cụ PivotChart. Hình 9 là biểu đồ so sánh diện tích thả nuôi mới với diện tích bị thiệt hại của tôm sú trên toàn tỉnh trong năm 2015-2016, được chi tiết hóa đến từng tháng sử dụng PivotChart của excel.

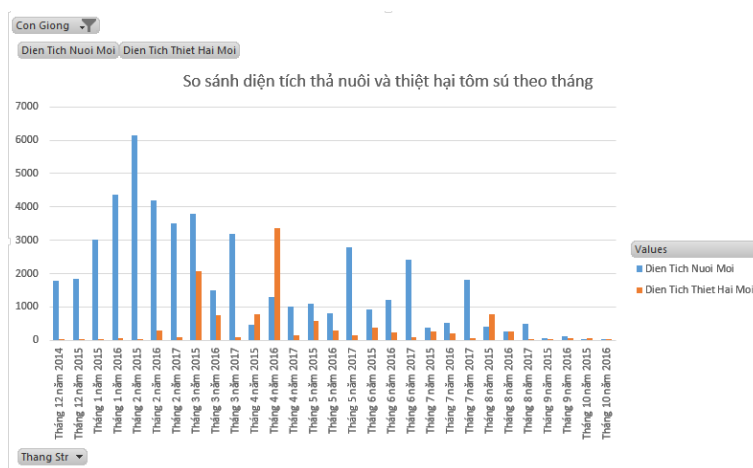
Để tăng tính kết nối, số liệu và biểu đồ báo cáo cũng cần được xuất bản lên web. Chúng tôi sử dụng dịch vụ báo cáo SQL Server Reporting Services 2017 để tạo một trang báo cáo trực quan, hỗ trợ hiển thị trên giao diện của điện thoại di động. Trang web báo cáo được thiết kế dưới dạng một bảng thông tin điều khiển (dashboard) để mỗi khi truy cập, cán bộ cấp quản lý có thể có ngay những báo cáo tổng quát về tình hình thả nuôi, thiệt hại ở các địa phương thể hiện dưới dạng những biểu đồ trực quan hoặc bảng số liệu. Hình 10 là một bảng thông tin thể hiện tình hình nuôi tôm sú từ đầu vụ 2016-2017, các biểu đồ thể hiện tình hình thả nuôi mới so sánh với tình hình thiệt hại, một bảng dữ liệu chi tiết cũng được thể hiện để chi tiết hóa các biểu đồ

Thao tác trực quan hóa số liệu nuôi trồng thủy sản từ dữ liệu tổng hợp kết xuất từ kho dữ liệu có ý nghĩa rất lớn, đặc biệt trong công tác cảnh báo dịch bệnh. Tùy thuộc nhu cầu nắm bắt thông tin của cán bộ quản lý, các biểu đồ khác nhau có thể được tạo ra để so sánh, thể hiện xu hướng, cảnh báo nguy hiểm đối với tình hình thả nuôi thủy sản trên địa bàn. Trong tình hình diễn biến phức tạp của các loại dịch bệnh thủy sản, việc hiển thị tỉ lệ thiệt hại so với thả nuôi của các loại thủy sản có giá trị cao như tôm sú, tôm chân trắng, cá tra, cá lóc... có ý nghĩa hết sức quan trọng. Các loại biểu đồ cảnh báo này giúp cho cán bộ quản lý nhanh chóng nắm bắt được tình hình chung về thiệt hại để có quyết định chiến lược hợp lí. Dịch vụ báo cáo SQL Server Reporting Services 2017 cũng có cơ chế phân quyền truy cập đối với các cấp độ người dùng khác nhau, nhờ đó, cơ chế phân cấp về truy cập, kết xuất báo cáo được đảm bảo.

Nhờ kết quả của các công cụ kết xuất này,

1	Thời Gian	2015-2016		
3	Row Labels	So Ho Nuoì Moi	Diện Tích Nuoì Moi	SL Giông Nuoì Moi
4	Tỉnh Trà Vinh	52056.9	38753.45	4638589.68
5	Huyện Càng Long	5523	998.05	27014
6	Huyện Cầu Kè	2	264	10558.4
7	Huyện Cầu Ngang	10546	4760.74	1679627.08
8	Cá các loại	820	84.65	10924.48
9	Cá lóc	12	10.7	2305
10	Cua biển	48	22.4	813
11	Nghêu	2	125	84.6
12	Tôm chân trắng	6101	2781.3	1332613
13	Tôm sú	3563	1736.69	332887
14	Huyện Châu Thành	8188.9	5265.71	540746.5
15	Huyện Duyên Hải	11819	15825.9	993592
16	Huyện Tiểu Cần	3034	607.5	32225.8
17	Huyện Trà Cú	4411	749.9	159369.9
18	Thành phố Trà Vinh	1	17	7500
19	Thị xã Duyên Hải	8532	10264.65	1187956
20	Grand Total	52056.9	38753.45	4638589.68

Hình 8: Kết xuất báo cáo dạng PivotTable từ kho dữ liệu bằng Microsoft Excel

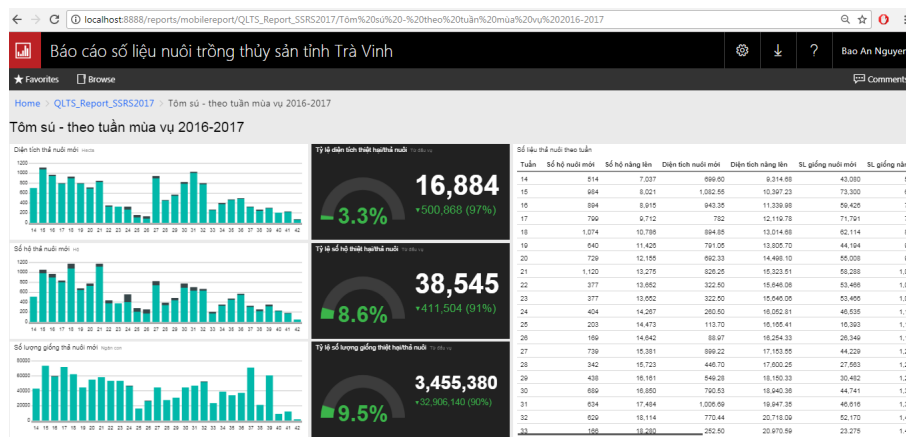


Hình 9: Kết xuất báo cáo dạng PivotChart từ kho dữ liệu bằng Microsoft Excel

cán bộ quản lí có thể dễ dàng nắm bắt tình hình chung, tổng hợp thông tin theo nhiều chiều với nhiều tiêu chí khác nhau, nhờ đó việc tạo lập báo cáo cũng được rút ngắn về thời gian và nâng cao về chất lượng. Bên cạnh đó, nếu được cập nhật tương đối đầy đủ số liệu, chúng ta có thể dễ dàng thực hiện các thao tác khai thác trên dữ liệu trong kho (tiên đoán, tìm luật kết hợp, phân lớp dữ liệu, gom cụm dữ liệu, phân tích tương quan...) để rút ra những quy luật, tương quan trong các biến động về tình hình nuôi trồng thủy sản, từ đó có thể có các dự báo để phục vụ cho các quyết định quản lí.

IV. KẾT LUẬN

Việc ứng dụng các công cụ trợ giúp ra quyết định đã được nghiên cứu và sử dụng rộng rãi ở các tổ chức doanh nghiệp. Bài báo này đề xuất một ứng dụng trực tuyến quản lí tập trung việc báo cáo số liệu nuôi trồng thủy sản trong tỉnh Trà Vinh. Nhờ đó, nó đảm bảo tính kịp thời, chính xác của công tác báo cáo. Bên cạnh đó, một kho dữ liệu sử dụng cơ sở dữ liệu đa chiều và các công cụ hỗ trợ cũng được phát triển nhằm trả lời các truy vấn mang tính tổng hợp phục vụ cho công tác quản lí. Nhờ tận dụng các tiện ích OLAP của excel, người dùng có thể dễ dàng thao tác truy vấn dữ liệu trên cơ sở dữ liệu đa chiều với excel mà không cần phải học sử dụng các



Hình 10: Báo cáo dạng dashboard sử dụng SQL Reporting Services 2017

công cụ phức tạp khác. Các tiện ích về trực quan hóa số liệu cũng được sử dụng để vẽ các biểu đồ và tạo một dashboard trực quan để hỗ trợ cán bộ quản lý ra quyết định. Mô hình hệ thống đơn giản, gọn nhẹ, tận dụng được các công nghệ được hỗ trợ trong hệ quản trị cơ sở dữ liệu Microsoft SQL Server nên dễ cài đặt, dễ triển khai và bảo trì. Để cung cấp thêm nhiều góc nhìn cho các cấp quản lý, trong tương lai, nếu thu thập đủ số liệu thì kho dữ liệu của hệ thống sẽ được cập nhật thêm chiều điều kiện tự nhiên, bao gồm các thông tin về thời tiết, độ mặn, mực nước... Nhờ đó, chúng ta có những phân tích dự báo gắn với điều kiện tự nhiên của mùa vụ nhằm hỗ trợ tốt hơn công tác quản lý nuôi trồng thủy sản trong tỉnh. Với các tiện ích đã nêu, mô hình hệ thống được đề xuất có thể được nhân rộng để ứng dụng trong các cơ quan nhà nước khác, nhằm đáp ứng nhu cầu quản lý ngày càng cao ở các lĩnh vực kinh tế xã hội trong tỉnh.

CẢM ƠN

(**) Nghiên cứu này được tài trợ từ nguồn kinh phí nghiên cứu khoa học của Trường Đại học Trà Vinh

TÀI LIỆU THAM KHẢO

- [1] William H Inmon. *Building the Data Warehouse*. 3rd ed. John Wiley & Sons, editor; 1992.
- [2] Jiawei Han, Micheline Kamber. *Data Mining: Concepts and Techniques*. 2nd ed.; 2012.
- [3] Surajit Chaudhuri, Umeshwar Dayal. An Overview of Data Warehousing and OLAP Technology. *ACM SIGMOD Record*. 1997;26(1):65–74.

- [4] Mickey Yost. Data Warehousing and Decision Support at the National Agricultural. *Journal Social Science Computer Review*. 2000;18(4):434–441.
- [5] Aditya Kumar Gupta, Bireshwar Dass Mazumdar. Multidimensional schema for agricultural data warehouse. *International Journal of Research in Engineering and Technology*. 2013;02(03).
- [6] Lê Quyết Thắng. *Nghiên cứu xây dựng các hệ thống thông tin hỗ trợ việc phòng chống dịch bệnh cây trồng và thủy sản cho vùng kinh tế trọng điểm*. Bộ Khoa học và Công nghệ, Hà Nội; 2010.
- [7] Phan Quốc Nghĩa, Nguyễn Nhật Lam. Ứng dụng công nghệ OLAP trong khai thác số liệu dịch hại trên lúa tại Trà Vinh. *Tạp chí Khoa học, Trường Đại học Trà Vinh*. 2012;6:7–15.
- [8] Trần Lê, Trương Chí Quang, Lê Văn Thạnh, Võ Quang Minh, Phạm Văn Quỳnh. Xây dựng hệ thống thông tin chuyên ngành nông nghiệp phát triển nông thôn trên WebGIS. *Tạp chí Khoa học, Đại học Cần Thơ*. 2013;Số Công nghệ Thông tin:1–10.
- [9] Phạm Thị Xuân Lộc. *Datamart và ứng dụng OLAP với kho dữ liệu lúa- tôm- cá ở Đồng bằng sông Cửu Long*. Hội thảo Quốc gia lần thứ XV "Một số vấn đề chọn lọc của Công nghệ thông tin và Truyền thông"; 2013.
- [10] Bộ Nông nghiệp và Phát triển nông thôn. *Dự án Phát triển Hệ thống Thông tin Quản lý ngành Lâm nghiệp tại Việt Nam - Giai đoạn II (FORMIS II)*; 2016.