

CÔNG THÔNG TIN QUẢN LÝ ĐÀO TẠO TÍCH HỢP DỊCH VỤ QUẢN LÝ ĐIỂM VÀ HỌC PHÍ SINH VIÊN TRƯỜNG ĐẠI HỌC CỬU LONG

Nguyễn Thị Hồng Nhung*

TÓM TẮT

Trong bài viết này, tôi nghiên cứu giải pháp cho vấn đề tích hợp dữ liệu từ nhiều nguồn tạo dữ liệu khác nhau bằng kỹ thuật tích hợp ETL trong Data Warehouse. Phát triển ứng dụng theo kiến trúc hướng dịch vụ, công nghệ Web Service, Metadata, sử dụng Framework Liferay Portal mã nguồn mở để xây dựng cổng thông tin quản lý đào tạo tích hợp dịch vụ quản lý điểm và học phí sinh viên trường Đại học Cửu Long.

Từ khóa: Kiến trúc hướng dịch vụ, dịch vụ web, Cổng thông tin Liferay, Kho dữ liệu, Siêu dữ liệu, Ngôn ngữ truy vấn SparQL

ABSTRACT

In this article, I study solutions to the problem of integrating data from many different data sources with integrated ETL specifications in the Data Warehouse. Application Development Service-oriented architecture, Web Service technology, Metadata, Liferay Portal Framework used to build open source portal management training service integration and management point of university student fees Mekong school.

Keywords: Service Oriented Architecture, Web Service, Liferay Portal, Data Warehouse, Metadata, SparQL

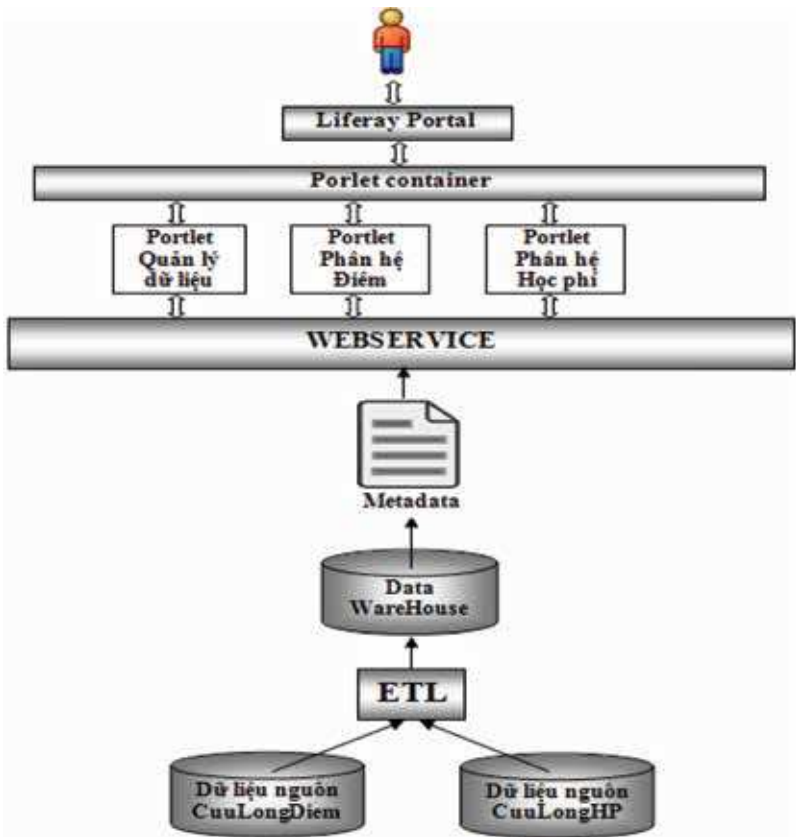
1. Đặt vấn đề

Trong những năm gần đây, công nghệ thông tin phát triển mạnh mẽ, phần mềm ngày càng được sử dụng rộng rãi trong mọi lĩnh vực của xã hội như doanh nghiệp, cơ quan, trường học,...Tuy nhiên, do xuất hiện nhiều công nghệ mới tạo nên môi trường không đồng nhất, nhiều hệ thống phần mềm quá phức tạp, gây khó khăn cho việc nâng cấp và bảo trì. Hệ thống phần mềm quản lý đào tạo của Đại học Cửu Long cũng gặp nhiều khó khăn như: các phần mềm đặt ở nhiều nơi, không liên kết, các yêu cầu chức năng thay đổi liên tục. Yêu cầu đặt ra là hệ thống phải có khả năng tích hợp, xử lý chính xác mà không ảnh hưởng đến dữ liệu gốc, có khả năng tái sử dụng, tiết kiệm chi phí,...Đồng thời tích hợp dữ liệu điểm và học phí trên cơ sở dữ

* Thạc sĩ, Khoa Công nghệ Thông tin, Trường Đại học Cửu Long

liệu đào tạo của trường Đại học Cửu Long để nâng cấp hệ thống hiện tại thành một hệ thống mới hoạt động hiệu quả hơn.

2. Mô hình công thông tin quản lý đào tạo tích hợp dịch vụ quản lý điểm và học phí trường Đại học Cửu Long



Hình 1. Mô hình tổng thể các công nghệ trong sử dụng trong hệ thống

3. Kho dữ liệu - Data Warehouse (DW)

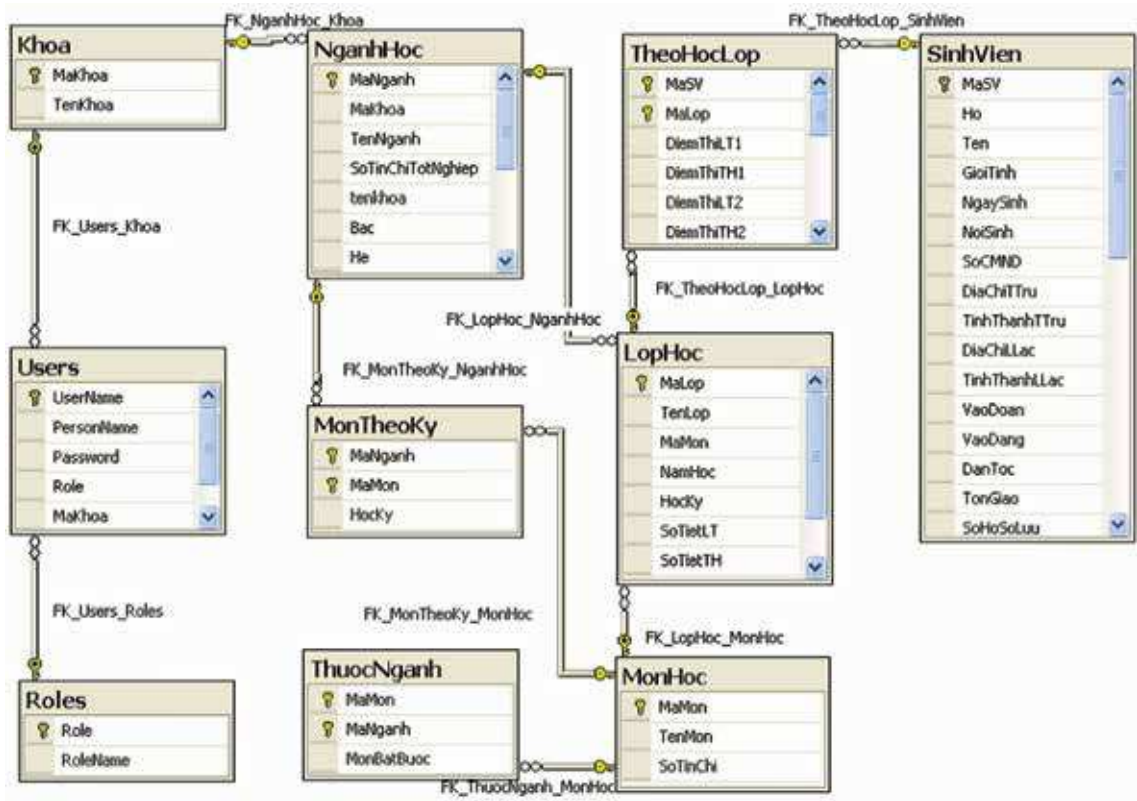
Kho dữ liệu: có khả năng thu thập dữ liệu từ nhiều nguồn dữ liệu hỗn tạp, đa bộ như cơ sở dữ liệu quan hệ, flat files và các bảng ghi giao dịch trực tuyến được trộn ghép với nhau tạo thành một thể thống nhất, cập nhật theo thời gian, được tích hợp theo hướng chủ đề nhằm hỗ trợ quá trình tạo quyết định về mặt quản lý.^[3]

Kho dữ liệu chủ đề (Datamart): có những đặc điểm giống kho dữ liệu nhưng với quy mô nhỏ hơn và lưu trữ dữ liệu về một lĩnh vực, một chuyên ngành. Datamart có thể được hình thành từ một tập con dữ liệu của kho dữ liệu hoặc có thể được xây dựng độc lập và sau khi xây dựng xong các Datamart có thể được kết nối, tích hợp lại với nhau tạo thành kho dữ liệu.^[3]

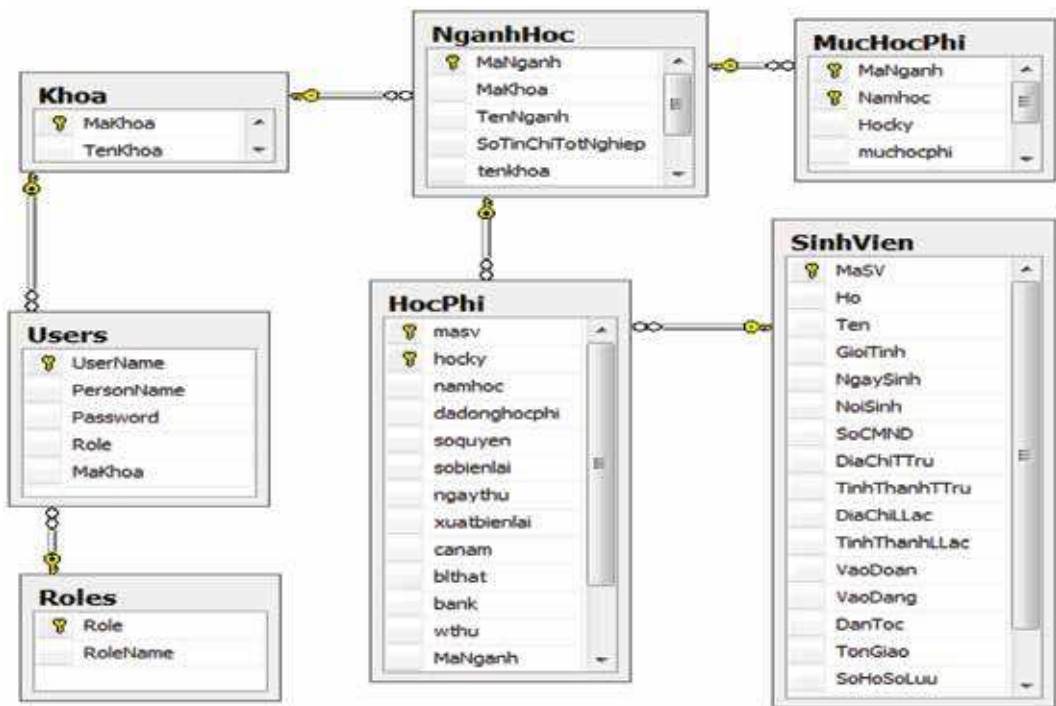
Xây dựng kho dữ liệu từ hai CSDL nguồn đã có: CSDL CuuLongDiem và CuuLongHP

Đánh giá CSDL nguồn: Do trường ĐH Cửu Long quản lý đào tạo theo niên chế nên mô

hình CSDL CuuLongDiem vẫn chưa được tối ưu. Thuộc tính MaLop sẽ lập đi lập lại nên dữ liệu bị thừa. Do đó cơ sở dữ liệu gốc cần thiết kế lại cho tối ưu hơn, tránh trùng lặp dữ liệu, hình 2, 3.



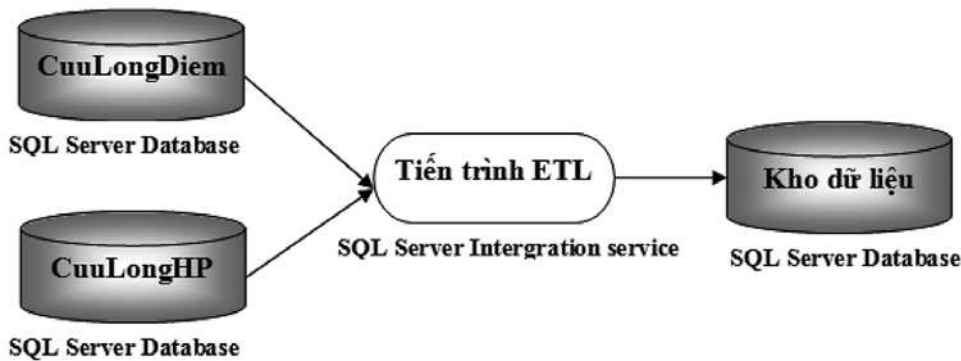
Hình 2. Mô hình CSDL CuuLongDiem



Hình 3. Mô hình CSDL CuuLongHP

Xây dựng tiến trình ETL

Dữ liệu cho hệ thống bao gồm dữ liệu quản lý điểm, dữ liệu quản lý học phí sinh viên. Dữ liệu này cần ánh xạ vào kho dữ liệu. Để thực hiện được ánh xạ này, tôi sẽ sử dụng công cụ SSIS để chuyển đổi dữ liệu từ hệ thống nguồn vào kho dữ liệu, hình 4.



Hình 4. Sơ đồ hoạt động ánh xạ dữ liệu nguồn vào kho dữ liệu

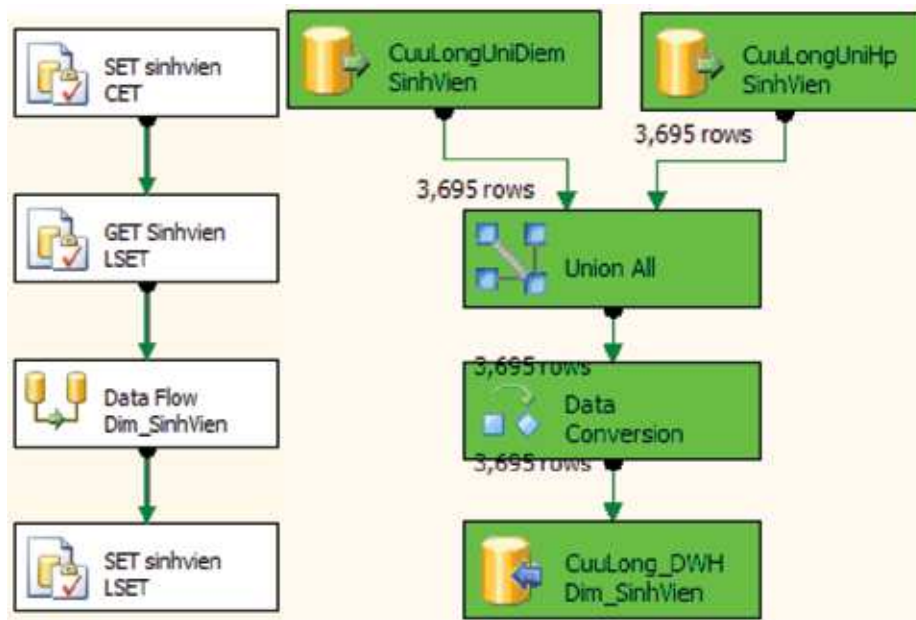
Các bước thực hiện: lấy dữ liệu từ nguồn, biến đổi dữ liệu, xây dựng các ánh xạ dữ liệu, thực thi các ánh xạ. Một số ánh xạ sẽ được thực thi định kỳ hằng ngày.

Ánh xạ dữ liệu cho các chiều: chiều sinh viên, chiều khoa, chiều ngành học, chiều môn học, chiều lớp học, chiều học phí, chiều mức học phí, bảng sự kiện fact_diem, bảng sự kiện fact_hocphi

Chiều Sinh viên

Dữ liệu của chiều Dim_SinhVien được ánh xạ từ hai bảng nguồn sinh viên trong cơ sở dữ liệu quan hệ CuuLongUniDiem và CuuLongUniHP. Dữ liệu được lấy bằng cách nối hai bảng lại với nhau, chuyển đổi kiểu dữ liệu với định dạng cho phù hợp với bảng chiều Dim_SinhVien trong kho dữ liệu. Sơ đồ ánh xạ như sau:

- *SET sinhvien*: Cập nhật thời điểm load dữ liệu vào kho
- *CET sinh viên LSET*: Lấy khoảng thời gian từ lần load cuối cùng đến thời điểm hiện tại
- *Dataflow*: quản lý luồng dữ liệu, thực hiện **ETL**
- *SET sinhvien LSET*: Cập nhật thời điểm load dữ liệu sau cùng vào kho
- *CuuLongUniDiem*: **E** - trích dữ liệu từ hệ thống nguồn kho nguồn
- *Union All*: **E** - Hợp nhất dữ liệu từ nhiều đầu vào thành một đầu ra
- *Data conversion*: **T** - đổi kiểu dữ liệu của các thuộc tính
- *CuuLong_DWH*: **L**- kết nối kho dữ liệu và nạp dữ liệu vào

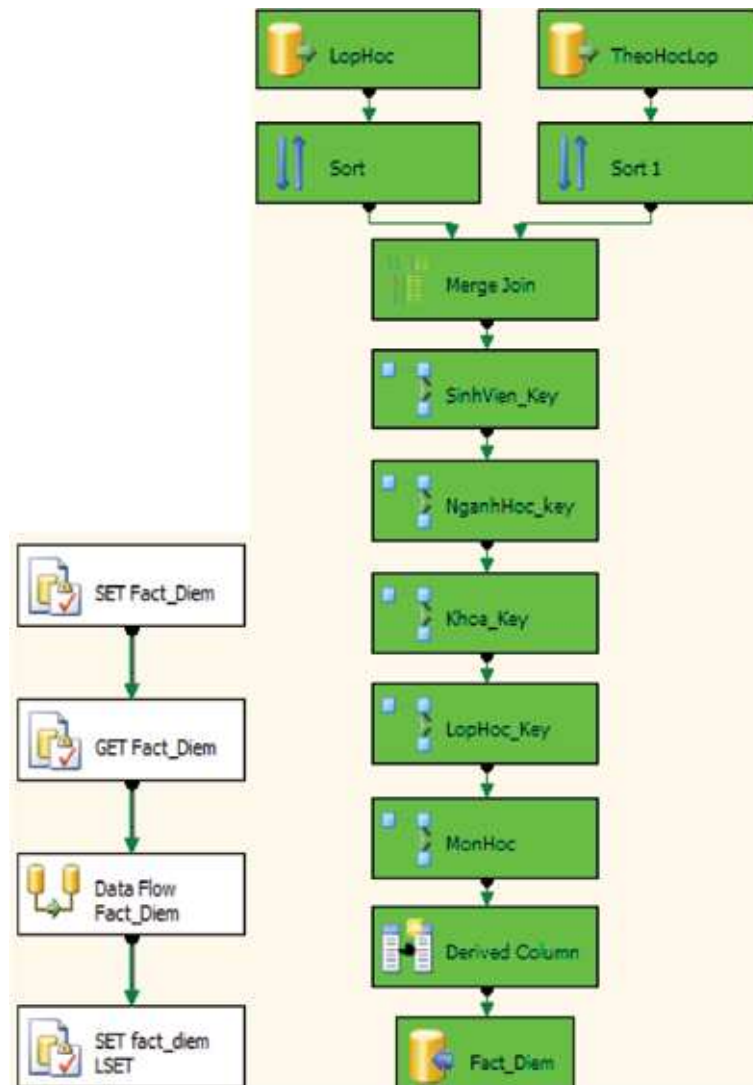


Hình 5. Sơ đồ ánh xạ chiều SinhVien

Bảng sự kiện *fact_Diem*

Bảng sự kiện Fact_Diem được tổng hợp dữ liệu từ các bảng LopHoc, TheoHocLop, Dim_SinhVien, Dim_Khoa, Dim_NganhHoc, Dim_MonHoc, Dim_LopHoc. Dữ liệu sau khi lấy ra sẽ được tổng hợp bằng các hàm tính toán và đưa vào kho dữ liệu.

- *SET fact_diem*: Cập nhật thời điểm load dữ liệu vào kho
- *GET fact_diem*: Lấy khoảng thời gian từ lần load cuối cùng đến thời điểm hiện tại
- *Dataflow*: quản lý luồng dữ liệu, thực hiện **ETL**
- *SET sinhvien LSET*: Cập nhật thời điểm load dữ liệu sau cùng vào kho
- *Sort*: sắp xếp dữ liệu trước khi hợp nhất dữ liệu
- *Merge join*: **E** – kết nối hai nguồn đã được sắp xếp trước
- *Lookup*: **T** – Thực hiện kết nối input với tập dữ liệu được tham chiếu
- *Derived Column*: **T**: xác định biểu thức để tạo ra cột mới hay xác định biểu thức để cập nhật column cũ.
- *Fact_diem*: **L**- kết nối kho dữ liệu và nạp dữ liệu vào



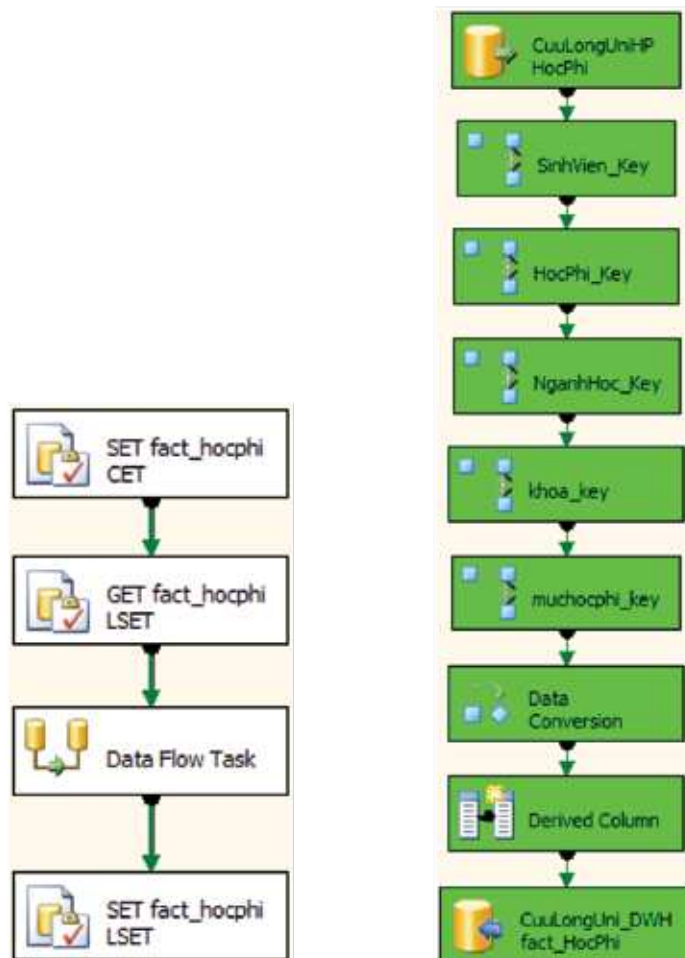
Hình 6. Sơ đồ ánh xạ fact_Diem

Bảng sự kiện fact_HocPhi

Bảng sự kiện Fact_HocPhi được tổng hợp dữ liệu từ các bảng HocPhi, Dim_SinhVien, Dim_Khoa, Dim_NganhHoc, Dim_HocPhi, Dim_MucHocPhi. Dữ liệu sau khi lấy ra sẽ được tổng hợp bằng các hàm tính toán và đẩy vào kho dữ liệu.

- *SET fact_diem*: Cập nhật thời điểm load dữ liệu vào kho
- *GET fact_diem*: Lấy khoảng thời gian từ lần load cuối cùng đến thời điểm hiện tại
- *Dataflow*: quản lý luồng dữ liệu, thực hiện **ETL**
- *SET sinhvien LSET*: Cập nhật thời điểm load dữ liệu sau cùng vào kho
- *Sort*: sắp xếp dữ liệu trước khi hợp nhất dữ liệu
- *Merge join*: E – kết nối hai nguồn đã được sắp xếp trước

- *Lookup: T* – Thực hiện kết nối input với tập dữ liệu được tham chiếu
- *Derived Column: T* - xác định biểu thức để tạo ra cột mới hay xác định biểu thức để cập nhật column cũ.
- *Fact_diem: L* - kết nối kho dữ liệu và nạp dữ liệu vào

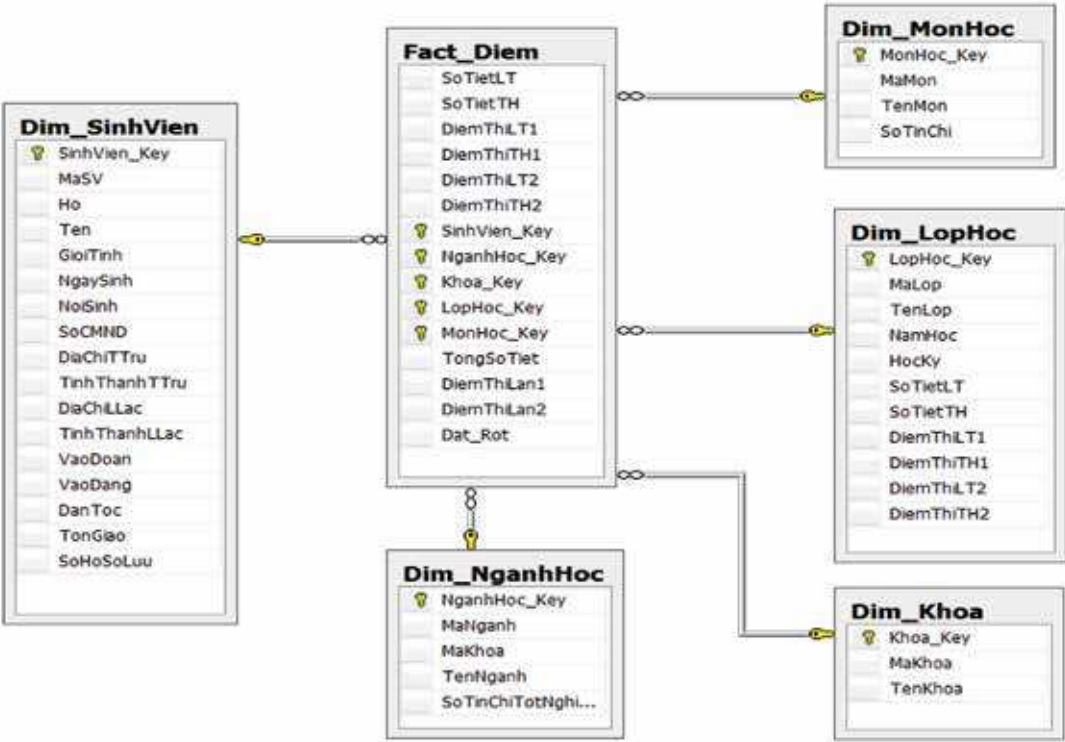


Hình 7. Sơ đồ ánh xạ fact_HocPhi

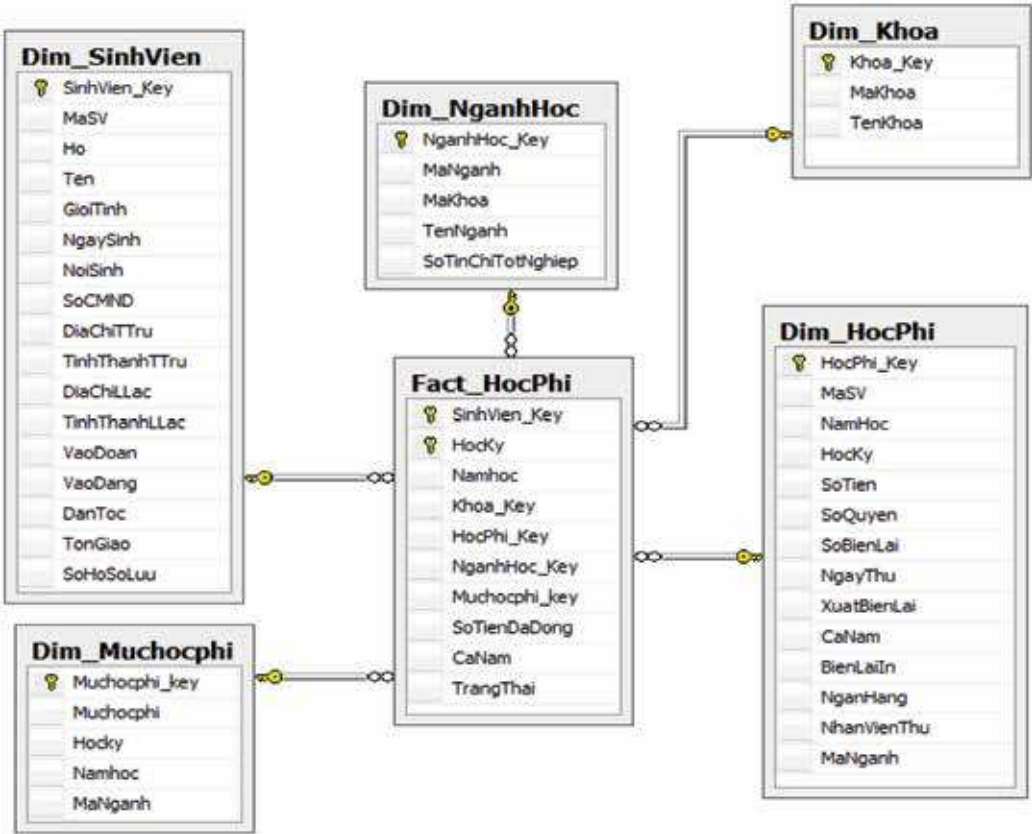
Kho dữ liệu được thiết kế gồm hai kho dữ liệu chủ đề: kho dữ liệu chủ đề điểm sinh viên và kho dữ liệu chủ đề học phí sinh viên

Kho dữ liệu chủ đề điểm sinh viên: là kho dữ liệu tổng hợp điểm sinh viên theo các tiêu chí sinh viên, khoa, ngành, lớp học và môn học. Các thông tin sự kiện bao gồm tổng số tiết học, điểm tổng hợp lần thi một, điểm tổng hợp lần thi hai, tình trạng kết quả đạt hay rớt của một sinh viên với một môn học.

Kho dữ liệu chủ đề học phí sinh viên: là kho dữ liệu lưu thông tin học phí sinh viên theo từng học kỳ của năm học. Các thông tin sự kiện gồm: số tiền đã đóng, đóng cả năm và trạng thái đã đóng hay chưa.



Hình 8. Mô hình kho dữ liệu chủ đề điểm sinh viên



Hình 9. Mô hình kho dữ liệu chủ đề học phí sinh viên

4. Kỹ thuật mô tả dữ liệu với RDF và Metadata

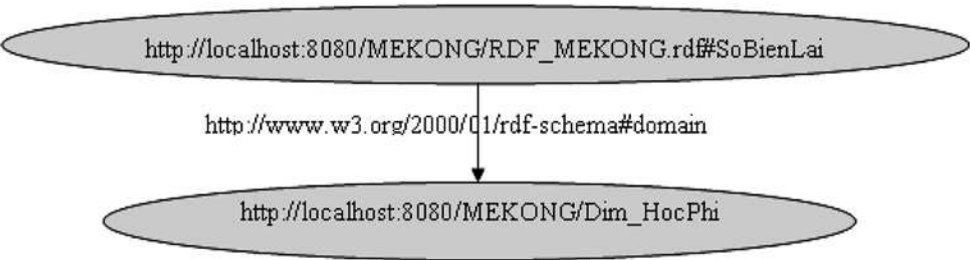
RDF là một “bộ khung” để mô tả các nguồn tài nguyên trên Internet, dựa trên cú pháp XML. RDF cung cấp một mô hình dữ liệu và một cú pháp đơn giản sao cho các hệ thống độc lập có thể trao đổi và sử dụng nó. RDF được xây dựng dựa trên cấu trúc bộ ba (subject - predicate – object) để thể hiện cho một phát biểu (statement). [2]

Ví dụ : triple A

Resource	<http://localhost:8080/MEKONG/RDF_MEKONG.rdf#SoBienLai>
Property	http://www.w3.org/2000/01/rdf-schema#label
Property value	Số Biên Lai

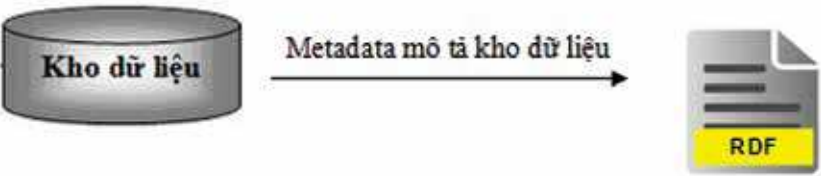
RDF Graph dùng để biểu diễn tập các triple. Các Subject và Object được biểu diễn là một node trong đồ thị và Predicate là một cung trong đồ thị, Cung của đồ thị luôn bắt đầu từ Subject đến Object. [2]

Ví dụ: Vẽ đồ thị cho phát biểu ở triple A



Hình 10. Đồ thị biểu diễn tập các triple

Metadata là dạng dữ liệu miêu tả về dữ liệu. Trong cơ sở dữ liệu quan hệ, metadata là các định nghĩa của bảng, cột, cơ sở dữ liệu, view và nhiều đối tượng khác. Trong kho dữ liệu, metadata là dạng định nghĩa dữ liệu như: bảng, cột, báo cáo, hay những quy tắc biến đổi. Metadata bao quát tất cả các phương diện của kho dữ liệu.[2]



Hình 11. Sơ đồ xây dựng file RDF mô tả kho dữ liệu

Xây dựng metadata cho kho dữ liệu

Từ mô hình kho dữ liệu, dùng metadata mô tả cho kho dữ liệu với cấu trúc XML

```

<?xml version="1.0" encoding="UTF-8" ?>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
- <rdfs:Class rdf:ID="Dim_Khoa">
  <rdfs:label>Chiều Khoa</rdfs:label>
  <rdfs:subClassOf rdf:resource="#DATAMARK_HOCPHI" />
</rdfs:Class>
- <rdf:Property rdf:ID="Khoa_key">
  <rdfs:label>Khoa Key</rdfs:label>
  <rdfs:domain rdf:resource="Dim_Khoa" />
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#bigint" />
</rdf:Property>
- <rdf:Property rdf:ID="MaKhoa">
  <rdfs:label>Mã Khoa</rdfs:label>
  <rdfs:domain rdf:resource="Dim_Khoa" />
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#nvarchar" />
</rdf:Property>
- <rdf:Property rdf:ID="TenKhoa">
  <rdfs:label>Tên Khoa</rdfs:label>
  <rdfs:domain rdf:resource="Dim_Khoa" />
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#nvarchar" />
</rdf:Property>
</rdf:RDF>

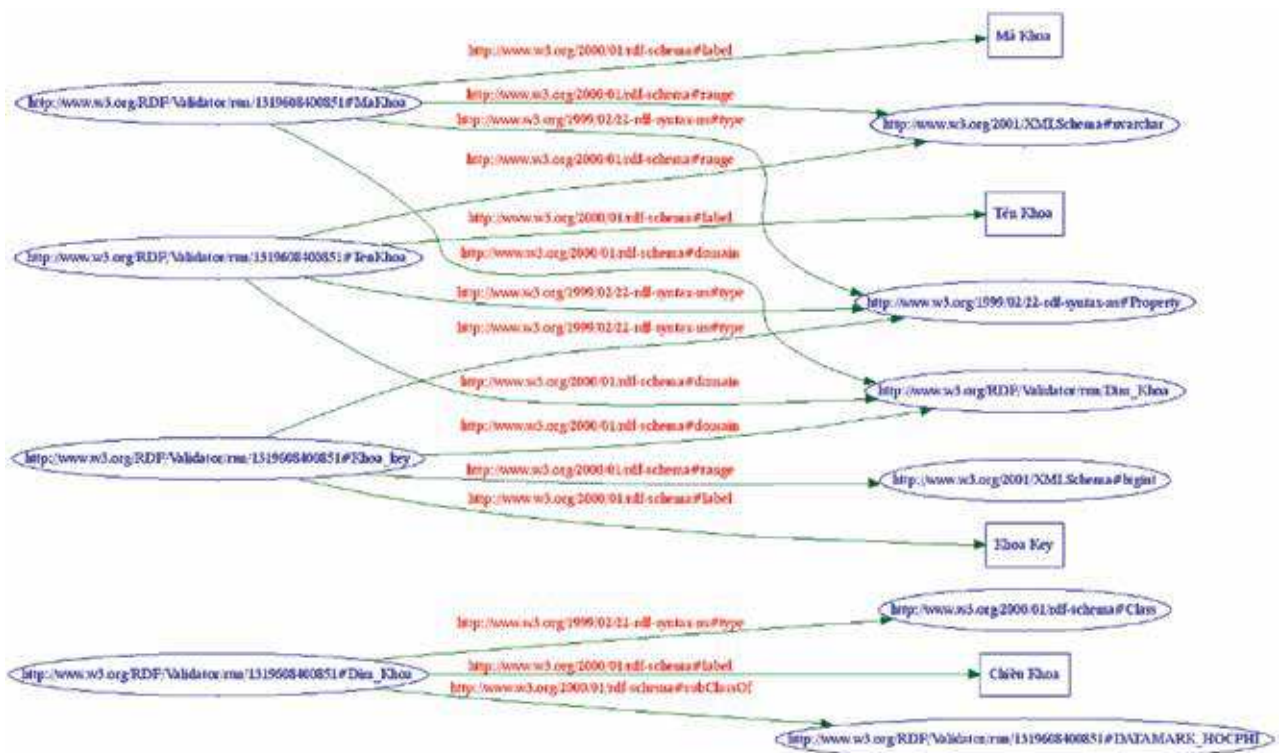
```

Hình 12. Một phần metadata trong file *RDF_DATA_DIEM.rdf*

Dùng chuẩn W3C RDF Validation Service” kiểm tra file RDF từ đường dẫn:

<http://www.w3.org/RDF/Validator/>

Kết quả kiểm tra file RDF là mô hình RDF graph như sau:



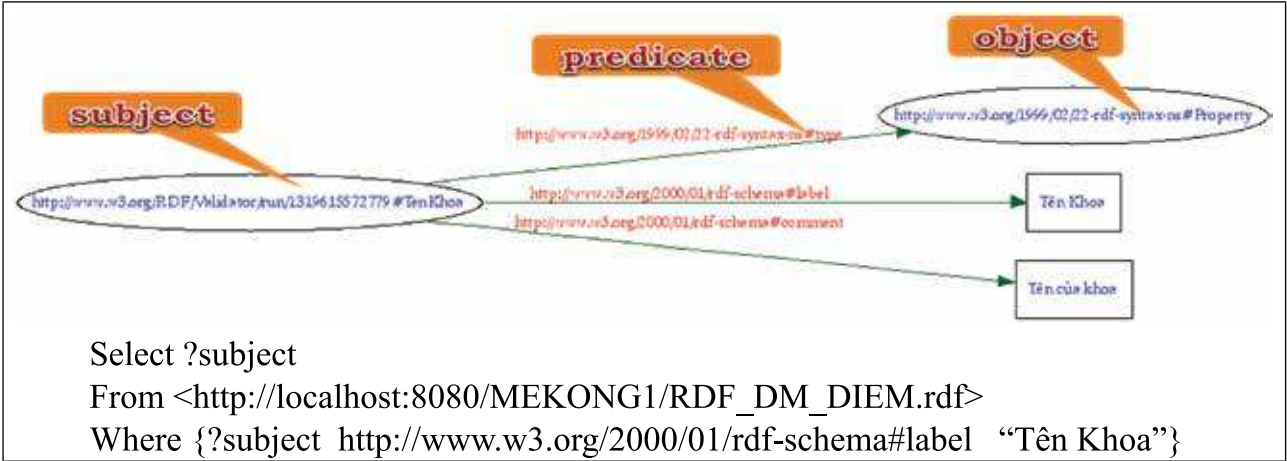
Hình 13. Một phần trong Mô hình RDF graph của kho dữ liệu

5. Ngôn ngữ truy vấn SparQL

SPARQL (Simple Protocol and RDF Query Language) là ngôn ngữ truy vấn RDF, được phát triển bởi tổ chức W3C được chuẩn hóa bởi nhóm phát triển RDF Data Access Working Group (DAWG).^[2]

Cấu trúc cơ bản của SparQL: gồm có hai mệnh đề, mệnh đề **SELECT** định danh các biến mà ứng dụng quan tâm, mệnh đề **WHERE** bao gồm các triple pattern. Một triple pattern là một RDF triple và các thành phần *subject*, *predicate*, *object* đều có thể thay thế là một biến.^[2]

– Ví dụ : có RDF graph sau



– Kết quả truy vấn

?subject
TenKhoa

6. Kiến trúc hướng dịch vụ và công nghệ dịch vụ web

SOA (Service oriented architecture): Một hệ thống được thiết kế theo kiểu hướng service là một hệ thống trong đó các chức năng của hệ thống được xây dựng dựa trên các service có độ kết dính thấp. Các service trong hệ thống giao tiếp với nhau thông qua việc gửi nhận các thông điệp (message).

Web Service là một công nghệ dùng để triển khai mô hình SOA vào thực tế. Nó đưa ra mô hình liên lạc trao đổi ứng dụng với ứng dụng trên ngôn ngữ đặc tả XML.

Để hệ thống có khả năng tái sử dụng các hàm xử lý cơ bản, hệ thống sẽ xây dựng theo kiến trúc hướng dịch vụ SOA, sử dụng công nghệ Web Service. Các thành phần xử lý chung của hệ thống sẽ được công bố vào môi trường mạng.

Gồm 4 gói dịch vụ:

– ***WS_DATA_DIEM, WS_DATA_HP***: lấy Metadata

- `get_table()`
- `get_attribule()`
- `get_label()`
- `get_att_of_label()`
- `get_table_of_label_att ()`
- `get_att_khoa ()`
- `get_att_nganh()`
- `get_att_mon()`
- `get_att_sinhvien()`
- `get_att_diem()`

– ***WS_DocDuLieu***: lấy dữ liệu

- `lay_ten_sv(masv)`
- `lay_tennganh_sv(masv)`
- `lay_hoc_ky_dang_hoc(masv)`
- `lay_ten_mon(monhoc_key)`
- `lay_DVHT(monhoc_key)`
- ...

– ***WS_XULY***

- `tongmondahoc_hocky(sinhvien, hk)`
- `tongmondatt_hocky(sinhvien, hk)`
- `tongsotinchidat_hocky`
- `tinhtb_nam(sv,hk)`

7. Cổng thông tin

Cổng thông tin điện tử tích hợp là điểm truy cập tập trung và duy nhất, tích hợp các kênh thông tin, các dịch vụ và ứng dụng, phân phối tới người sử dụng thông qua một phương thức thống nhất và đơn giản trên nền tảng Web. ^[1]

Liferay Portal là một trong những cổng thông tin mã nguồn mở hoàn chỉnh trên thị trường, cung cấp một khuôn khổ cơ bản để tạo ra bất kỳ loại cổng thông tin, hỗ trợ tất cả các máy chủ ứng dụng, cơ sở dữ liệu và hệ điều hành. ^[1]

Xây dựng cổng thông tin quản lý đào tạo trường Đại học Cửu Long

Quản lý xuất dữ liệu:

- Porlet quản lý xuất dữ liệu cho danh sách sinh viên rớt khóa
- Porlet quản lý xuất dữ liệu cho danh sách sinh viên đóng học phí
- Porlet quản lý xuất dữ liệu cho danh sách sinh viên được cấp học bổng
- Porlet quản lý xuất dữ liệu cho bảng điểm sinh viên

Phân hệ điểm:

- Porlet xuất danh sách sinh viên rớt khóa
- Porlet xuất danh sách sinh viên được *cấp học bổng*
- Porlet xuất bảng điểm sinh viên

Phân hệ học phí:

- Porlet xuất danh sách học phí sinh viên
- Porlet tính học phí cho sinh viên rớt khóa
- Porlet thống kê sinh viên đóng học phí



Hình 14. Porlet quản lý xuất dữ liệu cho danh sách sinh viên rớt khóa

8. Kết luận

Với mục tiêu nghiên cứu Data WareHouse để xây dựng kho dữ liệu tích hợp dữ liệu từ các cơ sở dữ liệu quản lý đào tạo đã có, nghiên cứu kiến trúc hướng dịch vụ để xây dựng các dịch vụ web xử lý, thống kê trên kho dữ liệu nhằm nâng cấp hệ thống phần mềm quản lý đào tạo hiện đang sử dụng tại trường Đại học Cửu Long sao cho hiệu quả hơn, tôi đã xây dựng hoàn chỉnh hệ thống ETL để xây dựng kho dữ liệu tích hợp cơ sở dữ liệu nguồn CuuLongDiem và CuuLongHP, ứng dụng công nghệ Web Service để xây dựng hệ thống xử lý điểm, học phí sinh viên, tách riêng dữ liệu và xử lý với Metadata sử dụng khung mô tả tài nguyên RDF, quản trị được portal trên hệ nền Liferay portal, xây dựng cổng thông tin quản lý đào tạo triển khai tích hợp dịch vụ điểm và học phí.

TÀI LIỆU THAM KHẢO

1. Dương Văn Ngân (2010), *Nghiên cứu cơ chế giao tiếp liên portlet trên hệ nền Liferay Portal*, Luận văn cao học, Khoa CNTT&TT, Đại học Cần Thơ, [9-48]
2. Nguyễn Thúc Anh Duy, Nguyễn Thị Khánh Hòa (2005), *Nghiên cứu tìm hiểu và xây dựng ứng dụng với Semantic Web*, Khóa luận cử nhân tin học, Khoa CNTT, Đại học Khoa học Tự nhiên, [35-97]
3. Dương Văn Ngân (2010), *Nghiên cứu cơ chế giao tiếp liên portlet trên hệ nền Liferay Portal*, Luận văn cao học, Khoa CNTT&TT, Đại học Cần Thơ, [9-48]
4. Vincent Rainardi (2008), *Building a Data Warehouse With Examples in SQL Server*, Apress, United States of America.
5. Ora Lassila, Ralph R. Swick (1999), *Resource Description Framework (RDF) Model and Syntax Specification*, Dan Brickley, University of Bristol

Ngày nhận bài: 27/12/2014

Ngày gửi phản biện: 25/6/2015