Thiết kế và triển khai kho dữ liệu khách hàng sử dụng dịch vụ viễn thông của Tổng Công ty Bưu chính Viễn thông Việt Nam (VNPT) : Luận văn ThS. Công nghệ thông tin: 1 01 10 /Nguyễn Văn Trung ; Nghd. : PGS.TS. Đặng Hữu Đạo

MỞ ĐẦU

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

Sự ra đời của công nghệ kho dữ liệu (Data Warehouse - DWH) trong những năm gần đây đã đáp ứng nhu cầu quản lý, lưu trữ một khối lượng dữ liệu 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.

Trong nước hiện nay có rất nhiều doanh nghiệp đã và đang nghiên cứu, tiến hành triển khai hệ thống DWH.

Đối với nước ngoài, hệ thống kho dữ liệu được áp dụng từ lâu và đã phát huy được những hiệu quả rất lớn giúp ích cho các doanh nghiệp trong việc hoạch định chiến lược kinh doanh cũng như việc nghiên cứu phát triển các ứng dụng.

Đối với Tổng công ty Bưu chĩnh Viễn thông Việt Nam (VNPT), xây dựng hệ thống Data Warehouse có ý nghĩa hết sức quan trọng. Ngoài việc thu thập lưu trữ các thông tin từ các hệ thống như Tính cước và chăm sóc khách hàng nói chung và các hệ thống khác nói riêng hệ thống DWH còn cung cấp các thông tin hữu ích giúp cho các nhà phát triển cập nhật các thay đổi hệ thống một cách nhanh chóng. Hơn thế nữa, hệ thống còn cung cấp cho nhà quản lý những thông tin quan trọng chính xác và nhanh chóng giúp họ có tầm nhìn chiến lược và hỗ trợ nhà quản lý ra những quyết định kịp thời và có lợi nhất cho doanh nghiệp trong điều kiện cạnh tranh của ngành viễn thông đang diễn ra rất gay gắt.

Xuất phát từ thực tế đó, đồ án tốt nghiệp này trình bày về việc ứng dụng công nghệ kho dữ liệu trên môi trường Oracle vào “Thiết kế và triển khai kho dữ liệu khách hàng sử dụng dịch vụ viễn thông của Tổng công ty Bưu chinh Viễn thông Việt Nam

(VNPT)" được thực hiện với mong muốn xây dựng được một DWH ban đầu để hỗ trợ VNPT đặc biệt là các Viễn thông tỉnh trong việc quản lý, điều hành doanh nghiệp.

Chương 1: TỔNG QUAN

1.1. Thị trường viễn thông

Trong những năm qua, thị trường viễn thông ở Việt Nam phát triển rất nhanh. Với chính sách mở cửa, khuyến khích cạnh tranh của Nhà nước, trên thị trường đã xuất hiện nhiều nhà cung cấp dịch vụ mới, cạnh tranh gay gắt với VNPT, khiến cho thị phần của VNPT bị chia sẻ đáng kể. Đặc biệt, trong điều kiện Việt Nam đã chính thức gia nhập WTO, thị trường viễn thông sẽ ngày càng sôi động hơn, VNPT sẽ phải đối mặt với sự cạnh tranh ngày càng gay gắt hơn không chỉ với các đối thủ trong nước mà còn với các đối thủ nước ngoài. Để có thể đứng vững và phát triển trong môi trường kinh doanh mới, VNPT cần phải không ngừng nâng cao năng lực cạnh tranh nói chung và đặc biệt là nâng cao năng lực cạnh tranh dịch vụ di động, băng thông rộng.

1.2. Dữ liệu ngành viễn thông

Ngành công nghiệp viễn thông lưu trữ một khối lượng dữ liệu khổng lồ, bao gồm chi tiết cuộc gọi, thông tin cảnh báo trình trạng của hệ thống mạng viễn thông và thông tin dữ liệu về khách hàng.

1.3. Mục tiêu luận văn

Xây dựng và triển khai DWH khách hàng sử dụng dịch vụ viễn thông cho Viễn thông tỉnh dựa trên việc tìm hiểu và phân tích hệ thống nguồn CCBS mà các viễn thông tỉnh đang sử dụng nhằm hỗ trợ lảnh đạo, các phòng ban,…đưa ra các chiến lược kinh doanh nhanh chóng chính xác. Xây dựng được CSDL tương đối hoàn chỉnh cho DWH; xây dựng các ánh xạ thu thập số liều từ hệ thống nguồn CCBS; xây dựng công cụ báo cáo, tra cứu và quản trị hệ thống; xây dựng được một số báo cáo phân tích ban đầu. Trong thiết kế hệ thống hướng tới tính mở và thiết kế tổng thể để dễ dàng mở rộng và áp dụng cho các doanh nghiệp viễn thông khác.

1.4. Phạm vi của luận văn

Xây dựng DWH đặc biệt là DWH viễn thông là một quá trình lâu dài với nhiều công việc. Trong luận văn này, dữ liệu DWH được xây dựng chủ yếu tập trung vào dữ liệu khách hàng và dữ liệu chi tiết cuộc gọi (Dữ liệu từ hệ thống nguồn CCBS). Quy mô hệ thống thử nghiệm được xây dựng áp dụng cho một Viễn Thông Tỉnh của Tập Đoàn Bưu Chính Viễn Thông Việt Nam VNPT. Mỗi Viễn Thông Tỉnh coi như mô hình thu nhỏ của VNPT, là nơi cung cấp đầy đủ các dịch vụ viễn thông tại địa bàn đó như dịch vụ: Điện thoại cố định, điện thoại di động, dịch vụ Gphone, điện thoại vệ

tinh, Internet, kênh thuê riêng,….Tập trung thiết kế mô hình dữ liệu và thu thập số liệu từ hệ thống nguồn CCBS, thiết kế một số các báo cáo, tra cứu cơ bản,

1.5. Công cụ thực hiện

- Hệ quản trị cở sở dữ liệu cho kho dữ liệu: Oralce.
- Công cụ xây dựng và quản trị kho dữ liệu: Oracle Warehouse Builder
- Công cụ xây dựng quản trị báo cáo:
  - Crystal Report thiết kế các mẫu báo cáo phân tích.
  - Crystal Report Server quản trị hệ thống báo cáo từ xa.
- Ngôn ngữ lập trình: Sử dụng ngôn ngữ C#, Asp.Net để xây dựng hệ thống báo cáo động và quản trị hệ thống kho dữ liệu.

1.6. Phương pháp xây dựng kho dữ liệu

Dựa trên cở sở của công nghệ phần mềm như mô hình thác nước để xây dựng DWH, nó bao gồm các bước chính sau: Nghiên cứu khả thi, yêu cầu, kiến trúc, thiết kế, phát triển, kiểm thử, triển khai, hoạt động. Có thể bổ sung vào mô hình thác nước hai bước thiết lập cơ sở hạ tầng và quản lý dự án.

Việc xây dựng kho dữ liệu thử nghiệm sử dụng Oracle Warehouse Builder có thể chia thành các giai đoạn sau:

- Phân tích yêu cầu.
- Thiết kế mô hình dữ liệu.
- Thiết kế vật lý.
- Thiết kế trích xuất chuyển đổi và nạp số liệu (ETL).
- Xây dựng hệ thống báo cáo, quản trị hệ thống.
- Cài đặt, triển khai và quản trị dữ liệu**.**

Chương 2: TỔNG QUAN VỀ DATA WAREHOUSE

2.1. Định nghĩa kho dữ liệu

Định nghĩa do W.H. Inman đề xướng: DWH được hiểu là một tập hợp các dữ liệu tương đối ổn định (không hay thay đổi), cập nhật theo thời gian, được tích hợp theo hướng chủ đề nhằm hỗ trợ quá trình tạo quyết định về mặt quản lý.

2.2. Đặc tính của kho dữ liệu

- Tính tích hợp.
- Gắn thời gian.
- Có tính lịch sử.
- Chỉ đọc

- Không biến động
- Dữ liệu tổng hợp và chi tiết

2.3. Kho dữ liệu và các cơ sở dữ liệu

DWH về bản chất cũng là một database bình thường, các hệ quản trị cơ sở dữ liệu quản lý và lưu trữ nó như các database thông thường tuy nhiên có hỗ trợ thêm về quản lý dữ liệu lớn và truy vấn. Thực ra nét khác biệt của DWH so với database là ở quan niệm, cách nhìn vấn đề.

2.4. Lợi ích của kho dữ liệu

- Tạo ra những quyết định có ảnh hưởng lớn.
- Công việc kinh doanh trở nên thông minh hơn.
- Dịch vụ khách hàng được nâng cao.
- Tái sáng tạo những tiến trình kinh doanh.
- Tái sáng tạo hệ thống thông tin.

2.5. Kho dữ liệu hiện nay

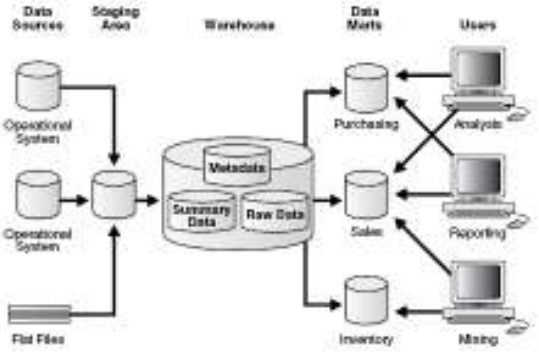
- Quản trị doanh nghiệp thông minh.
- Quản lý mối quan hệ khách hàng.
- Khai phá dữ liệu.
- Quản lý dữ liệu chủ.
- Tích hợp dữ liệu khách hàng.

2.6. Xu hướng tương lai của kho dữ liệu

- Dữ liệu phi cấu trúc.
- Tìm kiếm.
- Kiến trúc hướng dịch vụ.
- Kho dữ liệu thời gian thực.

2.7. Kiến trúc của kho dữ liệu

DWH và kiến trúc của nó tùy thuộc vào vị trí của từng tổ chức. Kiến trúc phổ biến của DWH:



Hình 4. Kiến trúc kho dữ liệu với Staging Area và Data Mart

2.7.1. Nguồn dữ liệu

Nguồn dữ liệu của DWH gồm nhiều loại khác nhau:

- Dữ liệu từ các hệ thống tác nghiệp.
- Hệ thống kế thừa.
- Các nguồn dữ liệu bên ngoài.

2.7.2. Công cụ trích xuất, chuyển đổi và nạp dữ liệu

2.7.3. Siêu dữ liệu

Lưu các định nghĩa logic các bảng, các thuộc tính của kho dữ liệu, tên các nguồn dữ liệu tác nghiệp, định nghĩa vật lý các bảng, các cột.

2.7.4. Kho dữ liệu chủ đề

Kho dữ liệu chủ đề (Data Mart - DM) là DWH có những đặc điểm giống với DWH nhưng với quy mô nhỏ hơn và lưu trữ dữ liệu về một lĩnh vực, một chuyên ngành. Các DM có thể được hình thành từ một tập con dữ liệu của kho dữ liệu hoặc cũng có thể được xây dựng độc lập và sau khi xây dựng xong, các DM có thể được kết nối tích hợp lại với nhau tạo thành kho dữ liệu. Vì vậy có thể xây dựng kho dữ liệu bắt đầu bằng việc xây dựng các DM hay ngược lại xây dựng kho dữ liệu trước sau đó tạo ra các DM.

2.7.5. Các công cụ truy vấn, tạo báo cáo, phân tích dữ liệu

- Công cụ tạo báo cáo và câu hỏi truy vấn (Report).
- Công cụ phân tích trực tuyến (OLAP).
- Công cụ phân tích, tìm kiếm cấp cao (Data Mining).

2.8. Tổ chức dữ liệu lôgíc

2.8.1. Lược đồ kho dữ liệu

a. Lược đồ hình sao

Trong sơ đồ hình sao, dữ liệu được xác định và phân loại theo 2 kiểu:

- Các sự kiện được tổ chức thành bảng Fact.

- Phạm vi, hay các chiều của dữ liệu, được tổ chức thành các bảng Dimension.

b. Lược đồ bông tuyết rơi

Sơ đồ hình tuyết rơi là một sự mở rộng của sơ đồ hình sao tại đó mỗi cánh sao không phải là một bảng Dimension mà là nhiều bảng.

c. Lược đồ kết hợp

Là kết hợp giữa sơ đồ hình sao và sơ đồ hình tuyết rơi. Một vài cơ sở dữ liệu và các công cụ truy vấn của người sử dụng đầu cuối nhất là các công cụ xử lí phân tích trực tuyến (OLAP) đòi hỏi mô hình dữ liệu phải là sơ đồ hình sao bởi vì nó là một mô hình dữ liệu quan hệ nhưng lại được thiết kế để hỗ trợ mô hình dữ liệu đa chiều là điểm cốt lõi của OLAP.

2.8.2. Mô hình dữ liệu đa chiều

Bản chất đa chiều của các câu hỏi trong nghiệp vụ được phản ánh trong thực tế chẳng hạn như những người quản lí thị trường không được thoả mãn với câu hỏi theo một chiều đơn giản, thay vào đó là những câu hỏi phức tạp. Một cách để quan sát một mô hình dữ liệu nhiều chiều là nhìn nó như một hình khối. Các thành phần chính:

- Các dữ kiện (Facts)
- Các chiều (Dimensions)
- Và các khối đa chiều (Cubes)

2.8.3. Bảng sự kiện

Bảng sự kiện điển hình có hai kiểu cột, chúng chứa đựng những sự kiện số (thường gọi là thước đo), và chứa khóa của các bảng dimension. Bảng sự kiện chứa đựng những sự kiện mức chi tiết hoặc những sự kiện mà đã được tổng hợp lại. Bảng sự kiện mà chứa sự kiện tổng hợp thường được gọi là những bảng tóm tắt. Bảng sự kiện thông thường chứa đựng những sự kiện với cùng mức của sự tổng hợp. Tuy nhiên hầu hết những sự kiện liên kết tất cả các chiều, nó có thể liên kết với 1 số chiều hoặc không liên kêt.

2.8.4. Bảng chiều

Các chiều là cách mô tả chủng loại mà theo đó các dữ liệu số trong khối được phân chia để phân tích. Khi xác định một chiều, chọn một hoặc nhiều cột của một trong các bảng liên kết (bảng chiều). Nếu ta chọn các cột phức tạp thì tất cả cần có quan hệ với nhau, chẳng hạn các giá trị của chúng có thể được tổ chức theo hệ thống phân cấp đơn. Để xác định hệ thống phân cấp, sắp xếp các cột từ chung nhất tới cụ thể nhất. Ví dụ: một chiều thời gian được tạo ra từ các cột năm, qúy, tháng, ngày.

Mỗi cột trong chiều góp phần vào một cấp độ cho chiều. Các cấp độ được sắp đặt theo nét riêng biệt và được tổ chức trong hệ thống cấp bậc mà nó thừa nhận các con đường hợp logic cho việc đào sâu.

2.8.5. Bảng sự kiện tổng hợp

Bảng tổng hợp nhằm mục đích trả lời nhanh các câu hỏi thường gặp.

2.9. Tổ chức dữ liệu vật lý

2.9.1. Phân vùng

Phân vùng (partition) là kĩ thuật được sử dụng trong kho dữ liệu nhằm tối ưu hiệu suất truy vấn bằng cách cho phép người thiết kế phân vùng các vùng nhớ để chứa dữ liệu thoả mãn những yêu cầu do người thiết kế đặt ra.

2.9.2. Chỉ mục

Đánh chỉ mục (Index) là kỹ thuật phổ biến nhằm tăng hiệu suất các truy vấn dữ liệu. Chuyên gia thiết kế sẽ chọn trường phù hợp của một bảng để đánh chỉ số, khi đó trường chỉ số đó sẽ được lưu ra một bảng tham chiếu, được sắp xếp sẵn. Khi có truy vấn dữ liệu, thời gian truy vấn sẽ giảm do dữ liệu cần truy vấn đã được sắp xếp từ trước.

Chương 3: PHÂN TÍCH, THIẾT KẾ VÀ TRIỂN KHAI KHO DỮ LIỆU KHÁCH HÀNG SỬ DỤNG DỊCH VỤ VIỄN THÔNG

3.1. Phân tích

Trong phần phân tích, tác giả sẽ tập trung vào 2 vấn đề chính:

- Tìm hiểu hệ thống nguồn CCBS đang được triển khai tại các Viễn thông tỉnh: Tìm hiểu hoạt động và cơ sở dữ liệu của hệ thống nguồn.
- Xác định các yêu cầu phần tích: Yêu cầu xuất phát từ nhu cầu cùa Viễn thông tỉnh và yêu cầu cần có của một DWH viễn thông.

3.1. 1. Tìm hiểu hệ thống nguồn CCBS

Hệ thống CCBS gồm 4 khối chức năng chính:

- Tính cước, chăm sóc và hỗ trợ khách. Khối này gồm các module:
  - Quản lý Phát triển thuê bao.
  - Tính cước
  - In hóa đơn
  - Quản lý nợ
  - Giải quyết khiếu nại
- Cắt mở thuê bao, dịch vụ tự động.
- Quản lý điều hành sửa chữa báo hỏng.

- Quản lý mạng ngoại vi.

3.1.2. Xác định yêu cầu phân tích

Các yêu cầu phân tích tập trung vào 6 nhóm chủ đề sau: bán hàng, quản lý mối quan hệ, kế toán, sử dụng dịch vụ, doanh thu và điều hành.

3.1.2.1. Bán hàng

- Phân tích chương trình khuyến mãi.
- Phân tích cung cấp sản phẩm chéo nhau
- Phân tích sự phát triển thuê bao
- Phân tích kênh bán hàng

3.1.2.2. Quản lý mối quan hệ

- Phân tích sự thay đổi nhà cung cấp dịch vụ.
- Phân tích hành vi của khách hàng.
- Phân tích sự khiếu nại của khách hàng.
- Phân tích khả năng tương tác với khách hàng.
- Phân tích tỷ lệ số tiền khách hàng sử dụng.
- Phân tích báo hỏng.
- Phân tích nhắc nợ cước.
- Phân tích khóa mở máy.
- Phân tích xử lý nợ cước.

3.1.2.3. Kế toán

- Phân tích nợ và thanh toán.
- Phân tích nợ không trả của khác hàng.

3.1.2.4. Phân tích cách sử dụng dịch vụ

3.1.2.5. Doanh thu

- Doanh thu từ khách hàng.
- Doanh thu của sản phẩm.

3.1.2.6. Điều hành

- Phân tích xử lý đơn hàng.
- Phân tích xử lý báo hỏng.

3.2. Thiết kế

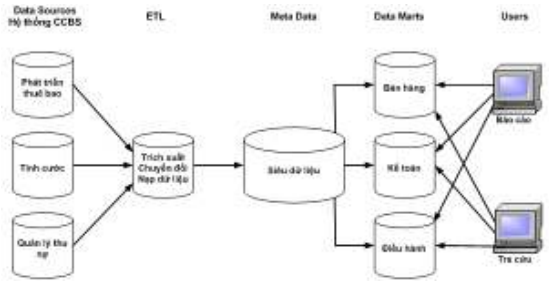
Thiết kế DWH sẽ tập trung vào các thiết kế chính sau:

- Thiết kế mô hình dữ liệu: Từ các yêu cầu cần phân tích ở trên và CSDL của hệ thống CCBS phân ra các chủ đề. Mỗi chủ đề có 1 mô hình dữ liệu tương ứng. Việc thiết kế mô hình dữ liệu qua các bước: chọn chủ đề, chọn các chiều, chọn các dữ kiện, xác định nguồn dữ liệu.

- Thiết kế mô hình vật lý:
  - Mô tả kiến trúc của DWH.
  - Từ dung lượng của hệ thống nguồn CCBS và các mô hình đã thiết kế ở phần thiết kế mô hình dữ liệu, ước lượng dung lượng của DWH.
  - Tạo cấu trúc CSDL của DWH từ các mô hình dữ liệu đã được thiết kế.
  - Từ yêu cầu cần phân tích và trả lời nhanh các câu hỏi, thiết kế một số bảng tổng hợp cần thiết.
  - Căn cứ dữ liệu của hệ thống nguồn CCBS, xác định các phân vùng và chỉ mục.
- Thiết kế trích xuất chuyển đổi nạp dữ liệu: Thiết kế các anh xạ dữ liệu từ hệ thống nguồn CCBS vào các mô hình dữ liệu đã thiết kế.
- Thiết kế công cụ báo cáo, tra cứu, quản trị hệ thống.

3.2.1. Kiến trúc của kho dữ liệu

Kiến trúc của DWH thử nghiệm bao gồm các thành phần chính: Dữ liệu nguồn CCBS, trích xuất chuyển đổi nạp dữ liệu, siêu dữ liệu, kho dữ liệu chủ đề, báo cáo và tra cứu như hình vẽ.



Hình 10. Kiến trúc của DWH thử nghiệm

- Hệ thống nguồn CCBS: Chứa dữ liệu khách hàng và chi tiết cuộc. Đây là dữ liệu nguồn của DWH. Hệ thống CCBS bao gồm các hệ thống con:
  - Quản lý phát triển thuê bao.
  - Tính cước.
  - In hóa đơn.
  - Quản lý nợ và xử lý nợ.
  - Giải quyết khiếu nại.
  - Quản lý điều hành sữa chữa báo hỏng.
- Trích xuất chuyển đổi nạp dữ liệu: Trích xuất các dữ liệu nguồn CCBS, chuyển đổi dữ liệu phù hợp khuôn dạng các kho dữ liệu chủ đề, nạp dữ liệu vào các kho dữ liệu chủ đề. Sử dụng OWB để xây dựng các ánh xạ trích xuất chuyển đổi và nạp dữ liệu. Các ánh xạ trích xuất chuyển đổi và nạp dữ liệu sẽ xây dựng:

- Ánh xạ cho các chiều như: khách hàng, thuê bao, thanh toán, địa chỉ, dịch vụ viễn thông, kiểu yêu cầu,….
- Ánh xạ cho các bảng sự kiện: bảng sự kiện bán hàng, phát triển thuê bao, điều hành thi công, khiếu nại, xử lý khiếu nại, báo hỏng, xử lý báo hỏng, chi tiết cước khách hàng, tổng hợp cước khách hàng, chi tiết nợ, chi tiết thanh toán, sử dụng dịch vụ gia tăng, sử dụng dịch vụ viễn thông,…

- Siêu dữ liệu: Lưu các định nghĩa logic các bảng, các thuộc tính của kho dữ liệu, tên các nguồn dữ liệu tác nghiệp, định nghĩa vật lý các bảng, các cột, các báo cáo, tra cứu, quản trị hệ thống,…
- Kho dữ liệu chủ đề: Có 16 kho dữ liệu chủ đề, 9 bảng tổng hợp sẽ được xây dựng nằm trong 6 chủ đề chính bán hàng, quản lý mối quan hệ, kế toán, cách sử dụng dịch vụ, doanh thu và điều hành.
  - Các kho dữ liệu chủ đề:
    - Bán hàng.
    - Phát triển thuê bao.
    - Điều hành thi công.
    - Khiếu nại.
    - Xử lý khiếu nại.
    - Báo hỏng.
    - Xử lý báo hỏng.
    - Chi tiết cước khách hàng.
    - Tổng hợp cước khách hàng.
    - Chi tiết nợ.
    - Chi tiết thanh toán.
    - Sử dụng dịch vụ gia tăng.
    - Sử dụng dịch vụ viễn thông.
    - Nhắc nợ.
    - Khóa mở máy.
    - Xử lý nợ.
  - Các bảng tổng hợp: Thống kê cuộc gọi theo giờ bắt đầu, thống kê cuộc gọi theo dịch vụ, thống kê cuộc gọi theo mã vùng, thống kê doanh thu, thống kê tiền nợ cước, thống kê tiền thanh toán, tổng hợp cước sử dụng, tổng hợp tiền nợ, tổng hợp tiền thanh toán.
- Báo cáo và tra cứu: Sử dụng công cụ xây dựng báo cáo và tra cứu động để thiết kế một số các báo cáo, tra cứu phân tích. Sử dụng công cụ tạo báo cáo và tra cứu động

dễ dàng tạo ra các báo cáo, tra cứu mới. Các báo cáo đã thiết kế đáp ứng các yêu cầu phân tích sau: Phân tích phát triển thuê bao, phân tích chiến dịch khuyến mãi, phân tích xử lý đơn hàng, phân tích khiếu nại, phân tích báo hỏng, phân tích doanh thu theo khách hàng, phân tích doanh thu theo sản phẩm, phân tích nợ, phân tích thanh toán,…Một số yêu cầu phân tích là sự tổng hợp của nhiều các thông tin như: phân tích sự dời bỏ nhà cung cấp, phân tích hành vi của khách hàng,…Danh sách các báo cáo đã xây dựng:

- Báo cáo phân tích bán hàng.
- Báo cáo phân tích chương trình khuyến mãi.
- Bảo cáo tổng hợp doanh thu theo thời gian.
- Báo cáo tổng thợp thu tiền theo thời gian.
- Báo cáo tổng thợp thanh toán theo thời gian.
- Thống kê danh sách khách hàng có doanh thu cao nhất.
- Thống kê danh sách khách hàng có doanh thu cao.
- Thống kê doanh thu theo sản phẩm.
- Báo cáo phân tích khiếu nại.
- Báo cáo phân tích báo hỏng.
- Thống kê cuộc gọi theo giờ bắt đầu.
- Thống kê cuộc gọi theo mã vùng.
- Thống kê cuộc gọi theo dịch vụ,…

3.2.2. Thiết kế mô hình dữ liệu

Từ các yêu cầu cần phân tích ở trên và CSDL của hệ thống CCBS phân ra 16 chủ đề sau: bán hàng, phát triển thuê bao, điều hành thi công, khiếu nại khách hàng, xử lý khiếu nại, cước chi tiết khách hàng, tổng hợp cước khách hàng, thanh toán nợ của khách hàng, nợ của khách hàng, nhắc nợ, khóa mở máy, xử lý nợ khó đòi, báo hỏng, xử lý báo hỏng, sử dụng dịch vụ viễn thông, sử dụng dịch vụ gia tăng.

Việc thiết kế mô hình dữ liệu qua các bước, từ yêu cầu ở trên chọn chủ đề cho mô hình, chọn các chiều, chọn các dữ kiện.

3.2.2.1. Bán hàng

Bán hàng ở đây chỉ xét đến bán sản phẩm dịch vụ viễn thông tức là khách hàng đến đăng ký sử dụng dịch vụ viễn thông như điện thoại cố định, di động, adsl,…và thay đổi các sử dụng dịch vụ viễn thông như di chuyển, thay đổi dịch vụ cộng thêm, thay đổi tốc độ,…Bán hàng xảy ra khi khách hàng đến đăng ký sử dụng dịch vụ viễn thông. Dữ liệu được lấy từ hệ thống phát triển thuê bao. Khách hàng đăng ký hằng ngày tại các kênh bán hàng của viễn thông tỉnh (điểm giao dịch, qua điện thoại, qua

internet,…). Các thông tin bán hàng như khách hàng, dịch vụ viễn thông, đơn vị tiếp nhận, thuê bao, kiểu yêu cầu, khuyến mãi, ngày yêu cầu, tiền đăng ký, vat đăng ký, tiền khuyến mãi, vat khuyến mãi,…Trong đó có thông tin của khóa ngoại của bảng khác hoặc dữ liệu sự kiện lấy từ hệ thống phát triển thuê bao hoặc thuộc tính suy dẫn cho việc phân tích. Sự kiện bán hàng đưa về bảng sự kiện như sau:

Các thông tin sự kiện mã thuê bao, tien_dk, vat_dk, tien_km, vat_km, ngay_tt, ngay_yc lấy từ dữ liệu phát triển thuê bao của hệ thống nguồn. Các thông tin như tien_tt, vat_tt được tính theo công thức tien_tt = tien_dk – tien_km, vat_tt = vat_dk – vat_km. Các thông tin khóa ngoại của bảng khác chính là các chiều của bảng sự kiện bán hàng như chiều khách hàng, dịch vụ viễn thông, kênh giao tiếp, điểm giao dịch, dịch vụ viễn thông, địa chỉ lắp đặt, chương trình khuyến mãi, thời gian.

3.2.2.2. Phát triển thuê bao

Phát triển thuê bao chỉ xét đến các thuê bao mà khách hàng đăng ký đã thi công xong và đưa vào sử dụng. Dữ liệu được lấy từ hệ thống phát triển thuê bao. Các thông tin của phát triển thuê bao như khách hàng, dịch vụ viễn thông, đơn vị tiếp nhận, thuê bao, kiểu yêu cầu, khuyến mãi, ngày yêu cầu, ngày sử dụng, ngày thanh toán,… Sự kiện kinh doanh đưa về bảng sự kiện như sau:

3.2.2.4. Khiếu nại của khác hàng

Khiếu nại lưu thông tin khiếu nại của khách hàng. Dữ liệu lấy từ hệ thống Giải quyết khiếu nại. Dữ liệu sử dụng phân tích tình hình khiếu nại của khác hàng theo

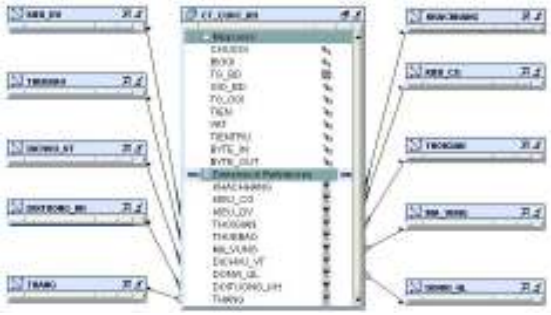
dịch vụ viễn thông, kênh tiếp nhận, nhóm khiếu nại,…Bảng sự kiện khiếu nại chứa các thông tin ngày khiếu nại, ngày hẹn trả lời, lần khiếu nại (lấy từ dữ liệu khiếu nại của hệ thống nguồn) và các trường khóa của các chiều khách hàng, thuê bao, dịch vụ viễn thông, nhóm khiếu nại, kênh giao tiếp, điểm giao dịch, thời gian.

3.2.2.5. Xử lý khiếu nại

Xử lý khiếu nại lưu thông tin xử lý khiếu nại của khách hàng. Dữ liệu lấy từ hệ thống Giải quyết khiếu nại. Dữ liệu sử dụng phân tích tình hình xử lý khiếu nại của khác hàng theo dịch vụ viễn thông, đơn vị xử lý, nhóm khiếu nại, kết quả khiếu nại,…Bảng sự kiện khiếu nại chứa các thông tin ngày khiếu nại, ngày hẹn trả lời, ngày xử lý, nội dung xử lý (lấy từ dữ liệu khiếu nại của hệ thống nguồn) và các trường khóa của các chiều khách hàng, thuê bao, dịch vụ viễn thông, nhóm khiếu nại, kết quả khiếu nại, đơn vị xử lý khiếu nại, thời gian.

3.2.2.6. Cước khách hàng

Cước khách hàng quản lý thông tin chi tiết cuộc gọi, truy nhập của khách hàng,…Mỗi cuộc gọi, truy nhập của khách hàng sinh ra một bản ghi chi tiết. Dữ liệu chi tiết cuộc gọi của khách hàng rất lớn, thường kết hợp các thông tin khác để phân tích hành vi của khách hàng, lưu lượng cuộc gọi theo tiêu chí khác nhau như mã vùng, kiểu cuộc gọi,….Dữ liệu được lấy từ dữ liệu cước chi tiết của hệ thống tính cước. Bảng sự kiện cước khách hàng như sau:



3.2.2.7. Tổng hợp cước khách hàng

Tổng hợp cước khách hàng quản lý thông tin cước sử dụng của khách hàng, được lưu chi tiết đến từng khoản mục, thuê bao, tháng. Dữ liệu tổng hợp sử dụng phân tích suất sinh lợi của khách hàng, sản phẩm,….Dữ liệu được lấy từ dữ liệu tổng hợp cước của hệ thống tính cước. Bảng sự kiện tổng hợp cước khách hàng như sau:

Các thông tin sự kiện như số cuộc, số phút, số giây, tiền, vat, tiền trừ,.. lấy từ dữ liệu tổng hợp cước của hệ thống nguồn. Các chiều của bảng sự kiện tổng hợp cước như chiều khách hàng, thanh toán, thuê bao, dịch vụ viễn thông, đơn vị quản lý, đối tượng khách hàng, tháng, khoản mục thanh cước.

3.2.2.8. Thanh toán nợ của khách hàng

Thanh toán nợ quản lý thông tin thanh toán nợ cước của khách hàng, được lưu chi tiết đến từng khoản mục, thuê bao, tháng. Dữ liệu tổng hợp sử dụng phân tích tình hình thanh toán của khách hàng. Dữ liệu được lấy từ dữ liệu thanh toán nợ cước của hệ thống quản lý thu nợ. Bảng sự kiện thanh toán nợ của khách hàng như sau:

3.2.2.9. Nợ cước của khách hàng

Nợ cước khách hàng quản lý thông tin nợ cước của khách hàng, được lưu chi tiết đến từng khoản mục, thuê bao, tháng. Dữ liệu nợ cước sử dụng phân tích tình hình nợ cước của khách hàng. Dữ liệu được lấy từ dữ liệu nợ cước của hệ thống quản lý thu nợ. Bảng sự kiện nợ cước của khách hàng như sau:

3.2.2.10. Nhắc nợ cước

Nhắc nợ cước quản lý thông tin nhắc nợ cước đến khách hàng. Dữ liệu nhắc nợ cước sử dụng phân tích tình hình nhắc nợ, đánh giá hiệu quả của quá trình nhắc nợ. Dữ liệu được lấy từ dữ liệu nhắc nợ cước của hệ thống quản lý thu nợ. Bảng sự kiện nhắc nợ cước của khách hàng như sau:

Các thông tin sự kiện như lần nhắc, ngày thực hiện,.. lấy từ dữ liệu nhắc nợ cước của hệ thống nguồn. Các chiều của bảng sự kiện nhắc nợ cước như chiều khách hàng, thuê bao, đơn vị quản lý, tháng, thời gian, hình thức nhắc nợ.

3.2.2.11. Khóa mở nợ cước

Khóa mở nợ cước quản lý thông tin khóa mở máy khi khách hàng không thanh toán tiền cước. Dữ liệu khóa mở nợ cước sử dụng phân tích tình hình khóa mở máy, đánh giá hiệu quả của quá trình khóa máy. Dữ liệu được lấy từ dữ liệu khóa mở máy của hệ thống quản lý thu nợ. Bảng sự kiện khóa mở như sau:

3.2.2.12. Xử lý nợ cước

Xử lý nợ cước quản lý thông tin xử lý nợ khó đòi khi khách hàng không thanh toán tiền cước. Dữ liệu xử lý nợ sử dụng phân tích tình hình xử lý nợ, đánh giá hiệu quả của quá trình xử lý nợ. Dữ liệu được lấy từ dữ liệu xử lý nợ của hệ thống quản lý thu nợ. Bảng sự kiện xử lý nợ như sau:

3.2.2.13. Báo hỏng

Báo hỏng quản lý thông tin khách hàng báo hỏng các dịch vụ viễn thông. Dữ liệu phân tích tình hình báo hỏng. Dữ liệu được lấy từ dữ liệu báo hỏng của hệ thống báo hỏng. Bảng sự kiện báo hỏng như sau:

3.2.2.14. Xử lý báo hỏng

Xử lý báo hỏng quản lý thông tin xử lý các báo hỏng của khách hàng. Dữ liệu phân tích tình hình xử lý báo hỏng. Dữ liệu được lấy từ dữ liệu xử lý báo hỏng của

hệ thống báo hỏng. Bảng sự kiện xử lý báo hỏng như sau:

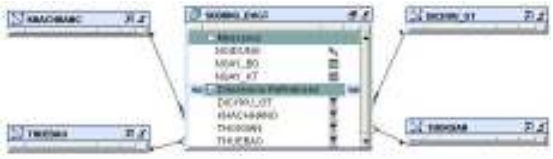
Các thông tin sự kiện như ngày báo hỏng, ngày xử lý, ghi chú xử lý,... lấy từ dữ liệu xử lý báo hỏng của hệ thống nguồn. Các chiều của bảng sự kiện xử lý báo hỏng như chiều khách hàng, thuê bao, đơn vị quản lý, thời gian, nguyên nhân, đơn vị thi công.

3.2.2.15. Sử dụng dịch vụ viễn thông

Sử dụng dịch vụ viễn thông lưu thông tin sử dụng dịch vụ viễn thông của khách hàng. Dữ liệu phân tích tình hình sử dụng dịch vụ vủa khách hàng như thế nào: dịch vụ được ai sử dụng, thời gian và ở địa chỉ nào. Dữ liệu được lấy từ dữ liệu quản lý khách hàng của hệ thống phát triển thuê bao. Bảng sự kiện sử dụng dịch vụ như sau:

3.2.2.16. Sử dụng dịch vụ gia tăng

Sử dụng dịch vụ viễn thông lưu thông tin sử dụng dịch vụ gia tăng của khách hàng. Dữ liệu phân tích tình hình sử dụng dịch vụ gia tăng vủa khách hàng như thế nào: dịch vụ gia tăng được ai sử dụng và thời gian sử dụng. Dữ liệu được lấy từ dữ liệu quản lý khách hàng của hệ thống phát triển thuê bao. Bảng sự kiện sử dụng dịch vụ gia tăng như sau:

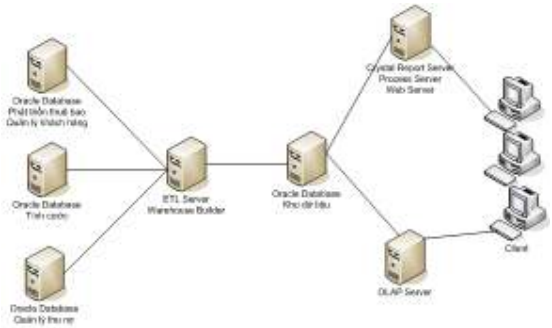


Các thông tin sự kiện như nội dung, ngày bắt đầu, ngày kết thúc,... lấy từ dữ liệu khách hàng của hệ thống nguồn. Các chiều của bảng sự kiện sử dụng dịch vụ như chiều khách hàng, thuê bao, thời gian, dịch vụ gia tăng.

3.2.2. Thiết kế mô hình vật lý

3.2.3.1. Kiến trúc vật lý kho dữ liệu

Kiến trúc hệ thống kho dữ liệu viễn thông bao gồm máy chủ chuyển đổi dữ liệu, máy chủ cở sở dữ liệu để lưu kho dữ liệu, máy chủ báo cáo + web server + process server, máy chủ phân tích dữ liệu trực tuyến.



Hình 53. Kiến trúc vật lý kho dữ liệu viễn thông

3.2.3.2. Ước lượng dung lượng

Xét tỉnh có số lượng thuê bao trung bình 250000 thuê bao, dữ liệu dự tính lưu trong 10 năm (khoảng 3650 ngày, 120 tháng).

Tổng dung lượng để lưu thông tinh khách hàng khoảng 400MB.

Tổng dung lượng để lưu chi tiết cuộc gọi trong 10 năm khoảng 480GB

Tổng dung lượng để lưu chi tiết cuộc gọi trong 10 năm khoảng 19GB

Chi tiết nợ của khách hàng trong trong 10 năm sử dụng 30GB; chi tiết trả của khách hàng trong trong 10 năm sử dụng 20GB; lịch sử khách hàng 10 năm sử dụng 20GB. Tổng các phần còn lại khoảng 100GB.

Như vậy tổng dung lượng để lưu trữ khoảng 800GB. Sử dụng thêm 30% để dánh chỉ mục (index), 15% cho cơ cấu bên trong, 20% cho sự sai số. Tổng dung lượng bây giờ sẽ là 1320GB

3.2.3.3. Tạo cấu trúc database

Trong phần thiết kế vật lôgic chúng ta đã thiết kế 18 bảng sự kiện, 28 chiều. Chúng ta sẽ tạo các bảng trong cở sở dữ liệu dựa trên thiết kế logic trên. Có thể sử dụng công cụ Warehouse Builder để thực hiện tạo bảng hoặc thực hiện theo các lệnh script.

3.2.3.4. Sử dụng bảng tổng hợp, view và materialized view

Như đã biết mục đích của các bảng tổng hợp trả lời nhanh các câu hỏi cần thường gặp. Các materialized view chính là các bảng tổng hợp (summary table), nó giống như view nhưng mang một số tính chất khác cơ bản: có chiếm bộ nhớ và do đó cũng có tốc độ thực hiện khá nhanh, quan hệ được với các dimension table và là những summary table được sử dụng để tăng tốc độ khai thác. Sau đây sẽ thiết kế một số bảng tổng hợp cho mục đích tổng hợp dữ liệu nhanh.

- ***Thống kê cuộc gọi theo giờ bắt đầu***

Bảng tổng hợp này được sử dụng để phân tích cuộc gọi theo giờ bắt đầu.

- ***Thống kê cuộc gọi theo dịch vụ***

Bảng tổng hợp này được sử dụng để phân tích cuộc gọi theo dịch vụ.

- ***Thống kê cuộc gọi theo mã vùng***

Bảng tổng hợp này được sử dụng để phân tích cuộc gọi theo mã vùng

- ***Thống kê doanh thu***

Bảng tổng hợp này được sử dụng để phân tích doanh thu theo dịch vụ, đối tượng, đơn vị quản lý, tháng.

- ***Thống kê tiền nợ cước***

Bảng tổng hợp này được sử dụng để phân tích tình hình nợ cước theo khoản mục, đơn vị quản lý, tháng.

- ***Thống kê tiền thanh toán***

Bảng tổng hợp này được sử dụng để phân tích tiền thanh toán nợ cước theo khoản mục, đơn vị quản lý, tháng, hình thức thanh toán.

- ***Tổng hợp cước sử dụng khách hàng***

Bảng tổng hợp này được sử dụng để phân tích cước sử dụng của khách hàng.

- ***Tổng hợp tiền nợ khách hàng***

Bảng tổng hợp này được sử dụng để phân tích tiền nợ của khách hàng.

- ***Tổng hợp tiền thanh toán khách hàng***

Bảng tổng hợp này được sử dụng để phân tích tiền thanh toán của khách hàng.

3.2.3.5. Phân vùng

Các partiton đóng vai trò quan trọng trong các bảng khai thác với số liệu vô cùng lớn. Nó cho phép phân chia dữ liệu theo một số tiêu thức nào đó. Trong kho dữ liệu viễn thông, chúng ta thiết kế các partition dựa trên các trường về thời gian. Dữ liệu chi tiết cuộc gọi, cước sử dụng, tiền thanh toán,…của khách hàng là rất lớn và thường được xử lý trong từng tháng, nên chúng ta sẽ phân vùng dữ liệu các dữ liệu này theo tháng. Danh sách các bảng được phân vùng dữ liệu

Tên bảng	**Trường index**
KHACHHANG	khachhang_id
THANHTOAN	thanhtoan_id
THUEBAO	thuebao_id
CT_CUOC_KH	khachhang
CT_CUOC_KH	thanhtoan
CT_CUOC_KH	thuebao
CT_NO_KH	khachhang
CT_NO_KH	thanhtoan
CT_TTNO_KH	khachhang
CT_TTNO_KH	thanhtoan
NHACNO	khachhang
NHACNO	thuebao
TH_CUOC_KH	khachhang
TH_CUOC_KH	thanhtoan
MV_TONGHOP_CUOC_SD_KH	khachhang
MV_TONGHOP_TIEN_NO_KH	khachhang
MV_TONGHOP_TIEN_TT_KH	khachhang

3.2.3.6. Chỉ mục

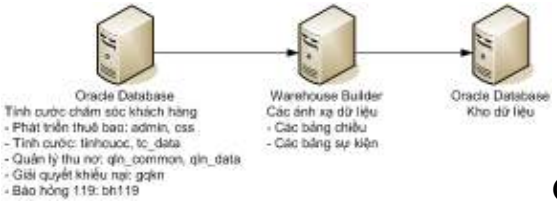
Thiết kế các chỉ mục (Index) làm tăng hiệu suất truy vấn dữ liệu. Chỉ nên đánh chỉ mục đối với các trường có sự khác nhau về dữ liệu nhiều. Sau đây liệt kê các bảng có thiết kế index.

Tên bảng	Trường phân vùng
CT_CUOC_KH	Tháng
CT_NO_KH	Tháng
CT_TTNO_KH	Tháng
NHACNO	Tháng
TH_CUOC_KH	Tháng
MV_TONGHOP_CUOC_SD_KH	Tháng
MV_TONGHOP_TIEN_NO_KH	Tháng
MV_TONGHOP_TIEN_TT_KH	Tháng

3.2.4. Thiết kế trích xuất chuyển đổi nạp dữ liệu

3.2.4.1. Mô tả chung

Hệ quản trị cơ sở dữ liệu hệ thống tính cước và chăm sóc khác hàng là oracle, gồm nhiều tài khoản dữ liệu như: Dữ liệu phát triển thuê bao bao gồm 2 user là admim, css; Dữ liệu cước khách hàng bao gồm 2 user tinhcuoc, tc_data; Dữ liệu quản lý thu nợ qln_common, qln_data; Dữ liệu giải quyết khiếu nại là user data gqkn; Dữ liệu báo hỏng 119 là user bh119,…Trong đề tài này sẽ sử dụng công cụ warehouser builder của oracle để thiết kế các ánh xạ dữ liệu từ hệ thống nguồn vào kho dữ liệu.



Các bước thực hiện: Tạo các kết nối vào cở sở dữ liệu nguồn, xây dựng các mapping dữ liệu, triển khai mapping, chạy mapping. Một số ánh xạ dữ liệu được chạy theo lịch định sẵn.

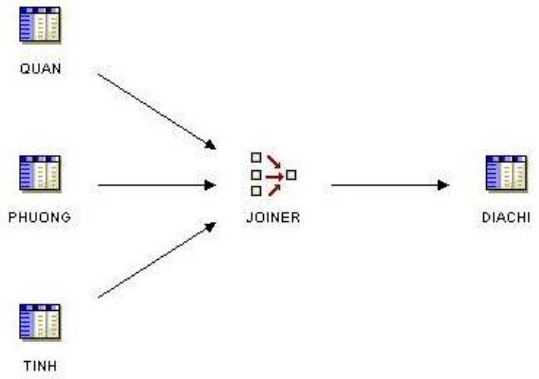
3.2.4.2. Ánh xạ cho các chiều

- ***Chiều sản phẩm (dịch vụ viễn thông)***

Chiều sản phẩm được lấy từ hai bảng DICHVU_VT và LOAIHINH_TB của dữ liệu phát triển thuê bao (user css). Sơ đồ ánh xạ DICHVU_VT_MAP cho chiều sản phẩm như sau:

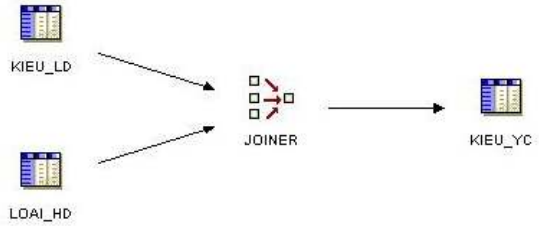
- ***Chiều địa chỉ***

Chiều địa chỉ được lấy từ các bảng TINH, QUAN, PHUONG của dữ liệu phát triển thuê bao (user css). Sơ đồ ánh xạ DIACHI_MAP cho chiều địa chỉ như sau:



- ***Chiều kiểu yêu cầu***

Chiều kiểu yêu cầu được lấy từ bảng LOAI_HD, KIEU_LD của dữ liệu phát triển thuê bao (user css). Sơ đồ ánh xạ KIEU_YC_MAP cho chiều kiểu yêu cầu như sau:



- ***Chiều khách hàng***

Chiều khách hàng được lấy từ bảng DB_KHACHHANG của dữ liệu phát triển thuê bao (user css). Sơ đồ ánh xạ KHACHHANG_MAP cho chiều khách hàng như sau:

- ***Chiều thanh toán***

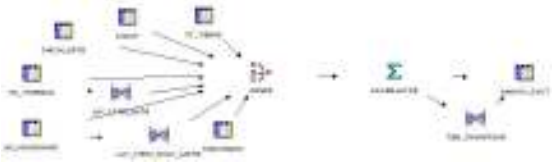
Chiều thanh toán được lấy từ bảng DB_THANHTOAN của dữ liệu phát triển thuê bao (user css). Sơ đồ ánh xạ THANHTOAN_MAP cho chiều thanh toán như sau:

- ***Chiều thuê bao***

Chiều thuê bao được lấy từ bảng DB_THUEBAO của dữ liệu phát triển thuê bao (user css). Sơ đồ ánh xạ THUEBAO_MAP cho chiều thuê bao như sau:

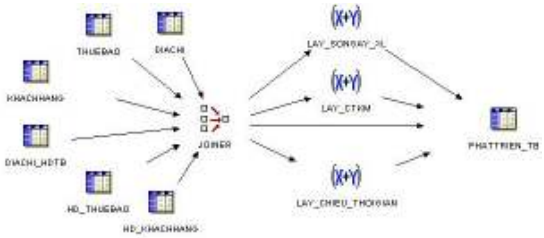
- ***Bảng sự kiện bán hàng (Đăng ký dịch vụ viễn thông)***

Bảng sự kiện bán hàng được lấy từ các bảng HD_KHACHHANG, HD_THUEBAO, DIACHI_HDTB, DIACHI, CT_TIENHD của dữ liệu phát triển thuê bao (user css). Sơ đồ ánh xạ DANGKY_DVVT_MAP cho bảng sự kiện bán hàng như sau:



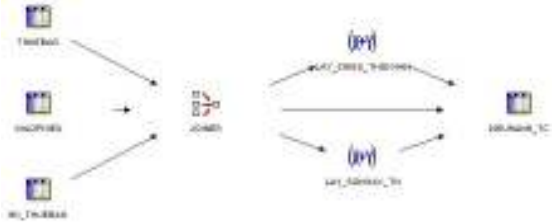
- ***Bảng sự kiện phát triển thuê bao***

Bảng sự kiện phát triển thuê bao được lấy từ các bảng HD_KHACHHANG, HD_THUEBAO, DIACHI_HDTB, DIACHI của dữ liệu phát triển thuê bao (user css). Sơ đồ ánh xạ PHATTRIEN_TB_MAP cho bảng sự kiện phát triển thuê bao như sau:



- ***Bảng sự kiện điều hành thi công***

Bảng sự kiện điều hành thi công được lấy từ các bảng GIAOPHIEU, HD_THUEBAO của dữ liệu phát triển thuê bao (user css). Sơ đồ ánh xạ DIEUHANH_TC_MAP cho bảng sự kiện điều hành thi công như sau:



- ***Bảng sự kiện khiếu nại***

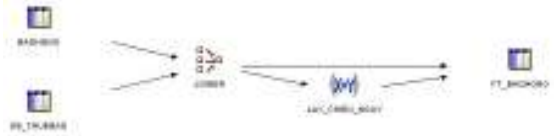
Bảng sự kiện khiếu nại lấy từ các bảng KHIEUNAI, CT_KHIEUNAI của dữ liệu khiếu nại (user qgkn), DB_THUEBAO (user css). Sơ đồ ánh xạ KHIEUNAI_MAP cho bảng sự kiện khiếu nại như sau:

- ***Bảng sự kiện xử lý khiếu nại***

Bảng sự kiện xử lý khiếu nại lấy từ các bảng KHIEUNAI, CT_KHIEUNAI, của dữ liệu xử lý khiếu nại (user qgkn), DB_THUEBAO(user css). Sơ đồ ánh xạ XL_KHIEUNAI_MAP cho bảng sự kiện xử lý khiếu nại như sau:

- ***Bảng sự kiện báo hỏng***

Bảng sự kiện báo hỏng lấy từ các bảng BAOHONG của dữ liệu báo hỏng (user bh119), DB_THUEBAO (user css). Sơ đồ ánh xạ BAOHONG_MAP cho bảng sự kiện báo hỏng như sau:

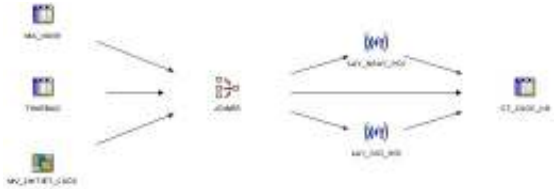


- ***Bảng sự kiện xử lý báo hỏng***

Bảng sự kiện xử lý báo hỏng lấy từ các bảng BAOHONG của dữ liệu xử lý báo hỏng (user bh119), DB_THUEBAO (user css). Sơ đồ ánh xạ XL_BAOHONG_MAP cho bảng sự kiện xử lý báo hỏng như sau:

- ***Bảng sự kiện chi tiết cước khách hàng***

Bảng sự kiện chi tiết cước khách hàng lấy từ các bảng MV_CHITIET_CUOC của dữ liệu tính cước (user tinhcuoc). Sơ đồ ánh xạ CT_CUOC_KH_MAP cho bảng sự kiện chi tiết cước khách hàng như sau:



- ***Bảng sự kiện tổng hợp cước khách hàng***

Bảng sự kiện tổng hợp cước khách hàng lấy từ các bảng MV_TONGHOP_CUOC của dữ liệu tính cước (user tinhcuoc). Sơ đồ ánh xạ TH_CUOC_KH_MAP cho bảng sự kiện tổng hợp cước khách hàng như sau:

- ***Bảng sự kiện chi tiết tiền nợ khách hàng***

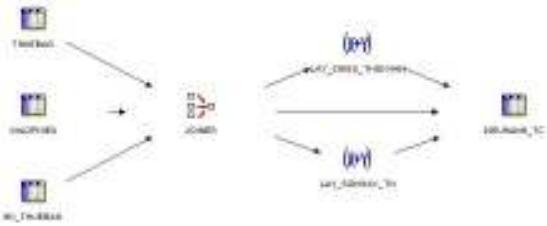
Bảng sự kiện chi tiết tiền nợ khách hàng lấy từ các bảng MV_NODAUKY của dữ liệu nợ đầu kỳ (user qln_data). Sơ đồ ánh xạ CT_NO_KH_MAP cho bảng sự kiện chi tiết tiền nợ khách hàng như sau:

- ***Bảng sự kiện chi tiết tiền thanh toán khách hàng***

Bảng sự kiện chi tiết tiền thanh toán của khách hàng lấy từ các bảng MV_CHITIET_TT_KH của dữ liệu thanh toán (user qln_data). Sơ đồ ánh xạ CT_TTNO_KH_MAP cho bảng sự kiện chi tiết tiền thanh toán của khách hàng như sau:

- ***Bảng sự kiện sử dụng dịch vụ viễn thông***

Bảng sự kiện sử dụng dịch vụ viễn thông được lấy từ các bảng DB_THUEBAO, DIACHI_TB, DIACHI của dữ liệu phát triển thuê bao (user css). Sơ đồ ánh xạ SUDUNG_DVVT_MAP cho bảng sự kiện sử dụng dịch vụ viễn thông như sau:



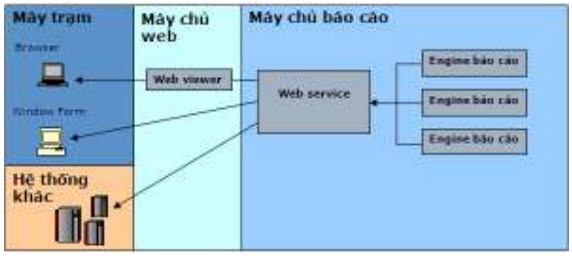
3.2.5. Thiết kế công cụ báo cáo, tra cứu động

Báo cáo, tra cứu là phần không thể thiếu trong các hệ thống kho dữ liệu. Việc xây dựng công cụ tạo báo cáo, tra cứu động và hệ thống báo cáo từ xa được xem xét

để phát triển hệ thống báo cáo, tra cứu. Công cụ cho phép quản lý báo cáo, tra cứu đầu ra và cho phép người dùng bổ xung báo cáo, tra cứu vào hệ thống, mở rộng khả năng khai thác của người dùng và đáp ứng nhu cầu sử dụng lâu dài.

3.2.5.1. Mô hình báo cáo từ xa

Mô hình báo cáo từ xa sử dụng Webservice được trình bày trên hình dưới. Trong mô hình này, engine báo cáo được chuyển lên máy chủ, và được máy chủ cung cấp giao diện ra bên ngoài thông qua Webservice.



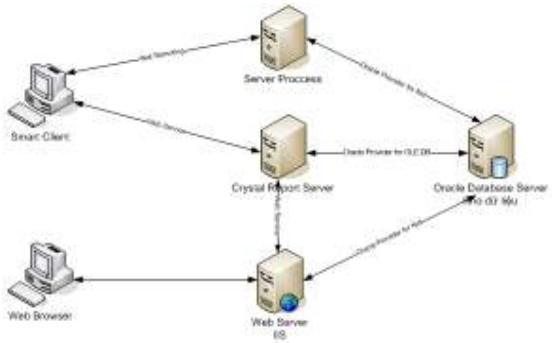
Hình 83. Mô hình báo cáo từ xa sử dụng Webservice

Việc chuyển engine báo cáo lên máy chủ sẽ giải quyết các nhược điểm của mô hình truyền thống. Trong khi đó, việc sử dụng Webservice mở ra các khả năng:

- Sử dụng thống nhất báo cáo cho ứng dụng web và ứng dụng Smart Client.
- Tích hợp dễ dàng hệ thống báo cáo tới các hệ thống khác thông qua môi trường Internet hoặc mạng LAN.

3.2.5.2. Kiến trúc và công nghệ

Với mô hình báo cáo từ xa lựa chọn ở trên kiến trúc của hệ thống báo cáo, tra cứu được xây dựng như sau.



Hình 94. Kiến trúc và công nghệ hệ thống báo cáo, tra cứu

Máy chủ báo cáo được sử dụng ở đây là Crystal Report Server, truy suất vào kho dữ liệu Oralce Database qua giao thức Oracle Provider OLE DB. Server Proccess, Web Server có chứa các ứng dụng viết bằng ngôn ngữ C#, ASP.Net kết nối với cở sở dữ liệu Oracle qua giao thức Oracle Provider Net.

3.2.5.3. Các chức năng cho hệ thống báo cáo

Có 2 nhóm chức năng để có thể tạo và xem báo cáo: nhóm chức năng định

nghĩa báo cáo, nhóm chức năng xem báo cáo.

a. Nhóm chức năng định nghĩa báo cáo

Chức năng định nghĩa báo cáo sẽ cho phép tạo ra báo cáo mới, phân quyền xem báo cáo mới cho người dùng hoặc nhóm người dùng. Các tham số, điều kiện của báo cáo này sẽ được lưu vào CSDL.

Các bước và quy tắc định nghĩa một báo cáo mới:

- Định nghĩa một danh mục báo cáo mới:

Nếu báo cáo cần hiển thị dưới dạng crystal report người dùng sẽ sử dụng Crystal Report Design để thiết kết mẫu báo và copy mẫu báo cáo lên Crystal Report Server.

Mỗi báo cáo, tra cứu có 1 câu lệnh để lấy dữ liệu. Trong câu lệnh đó có thể có các tham số truyền vào, cú pháp của tham số truyền vào là **{?ten_tham_so}**. Ví dụ báo cáo thống kê doanh thu, câu lệnh lấy dữ liệu có 5 tham số: **vttinh_id, nam, quy, thang, DK_DONVI_QL**.

```
SELECT ROW_NUMBER() OVER (ORDER BY a.donvi_ql desc)  stt,
c.ten_dv, sum(a.socuoc) cuoc, sum(a.tien) tien,
sum(a.vat) vat, sum(a.tong) tong
FROM wh_ccbs.mv_thongke_doanhthu a, wh_ccbs.thang b,
wh_ccbs.donvi_ql c
where a.thang = b.dimension_key
    and a.donvi_ql = c.donvi_id
    and a.donvi_ql in ({?DK_DONVI_QL})
    and c.vttinh_id = {?vttinh_id}
    and b.nam = {?nam}
    and decode({?quy},0,0,b.quy) = {?quy}
    and decode({?thang},0,0,b.thang) = {?thang}
group by a.donvi_ql, c.ten_dv
```

Trong chức năng định nghĩa báo cáo, tra cứu mới: Mỗi báo cáo có 1 câu lệnh SQL, tên file cystal report (Nếu muốn hiển thị kết quả báo cáo theo thiết kế của file crystal report), trạng thái của báo cáo, nhóm báo cáo,…

- Định nghĩa tham số cho báo cáo:

Mỗi báo cáo, tra cứu có thể có nhiều tham số. Có hai loại tham số: tham số chọn trên chức năng xem báo cáo, tra cứu và tham số phân quyền cho người dùng. Cú pháp của tham số phân quyền cho người sử dụng **{?DK_TENBANG},** chương trình sẽ thay thế tham số đó bằng danh sách giá trị của bảng (chiều) **TENBANG** mà

người dùng có quyền thao tác, loại tham số này không cần định nghĩa. Trong ví dụ trên tham số **DK_DONVI_QL** sẽ được chương trình thay thế bằng danh sách các đơn vị quản lý mà người dùng có quyền thao tác. Các tham số còn lại **vttinh_id, nam, quy, thang** là tham số được chọn trên chức năng xem báo cáo các tham số này phải được định nghía.

Trong chức năng định nghĩa tham số báo cáo mỗi tham số có các thuộc như sau:

- Tên tham số: Đặt tên giống với tên trong câu lệnh lấy dữ liệu báo cáo. Ví dụ: **vttinh_id**.
- Tên hiển thị: Tên hiển thị trên chức năng xem báo cáo, tra cứu.
- Loại điều khiển: Có các loại điều khiển như **Combobox, Textbox, Listbox, Checkbox, Ngay gio, Ngay, Thang**.
- Định dạng: Đối với loại điều khiển **Ngay gio, Ngay, Thang** phải có định dạng của dữ liệu, ví dụ tham số **thang** có định dạng **yyyyMM**.
- Lệnh: Đối với loại điều khiển **Combobox, Listbox** phải có kèm theo câu lệnh SQL lấy dữ liệu cho điều khiển đó, ví dụ tham số **vttinh_id** có câu lệnh lấy dữ liệu: Cột thứ nhất là id, cột thứ hai hiển thị trên giao diện.

```
select distinct a.vttinh_id, a.vt_tinh
from wh_ccbs.donvi_ql a
```

Các điều khiển Combobox có thể có quan hệ với nhau, Combobox trên giới hạn dữ liệu cho Combobox dưới. Ví dụ khi chọn tham số **quy**, để tham số **thang** chỉ hiển thị các tháng thuộc quý đó thì câu lệnh lấy dữ liệu cho tháng phải có cấu trúc:

```
SELECT distinct a.thang, substr(a.thang,5,2) aa
FROM wh_ccbs.thang a
where a.quy = :quy
order by thang
```

- Kiểu dữ liệu, có 2 loại: varchar và number.
- Vị trí: Thứ tự hiển thị điều khiển trên chức năng xem báo cáo, tra cứu.
- Giá trị text: Đối với điều khiển dạng **Combobox,** nếu có giá trị này thì tham số báo cáo được thay thế bằng giá trị text của **Combobox** (cột thứ 2 của câu lệnh lấy dữ liệu) ngược lại lấy theo id (cột thứ 1 của lệnh lấy dữ liệu).
- Hiển thị nhãn: Nếu có giá trị thì, sẽ có 1 tham số có tên **lblTenThamSo** tương ứng với tham số **TenThamSo**, giá trị của nó là giá trị text của điều khiển. Sử dụng tham số trong lệnh lấy dữ liệu hoặc hiển thị trên form kết quả.
- Định dạng nhãn: Định dạng cho tham số nhãn.

• Định nghĩa tham số cho báo cáo: Một báo cáo có thể nhiều báo cáo con. Mỗi báo

cáo con có câu lệnh tương ứng.

- Gán quyền xem báo cáo cho người dùng, nhóm người dùng: Sau khi định nghĩa xong báo cáo có thể gán báo cáo đó cho 1 quyền nào đó, quyền đó được gán cho người dùng hoặc nhóm người dùng.

b. Nhóm chức năng xem báo cáo, tra cứu

Hiển thị các báo cáo, tra cứu mà người dùng có quyền xem. Đối với từng báo cáo, tra cứu hiển thị các điều kiện để người dùng lựa chọn. Đối với báo cáo truyền yêu cầu báo cáo cho Crystal Report Server xử lý và hiển thị kết quả báo cáo cho người sử dụng. Đối với tra cứu truyền yêu cầu báo cáo cho Server Proccess xử lý và hiển thị kết quả tra cứu cho người sử dụng.

Đối với từng báo cáo, chương trình sẽ lấy danh sách các tham số trong CSDL. Đối với từng tham số, dựa vào loại điều khiện và thông số cho điều khiển đó, chương trình tự động thêm control vào vùng điều kiện trên chức năng hiển thị báo cáo, tra cứu. Sau khi người dùng chọn điều kiện xong và chấp nhận xem báo cáo tra cứu, chương trình lần lượt đọc các điều khiển, lấy giá trị người dùng đã chọn thay thế các tham số trong câu lệnh SQL lấy dữ liệu. Với câu lệnh SQL đã được xử lý, Crystal Report Server hoặc Server Proccess sẽ trả kết quả cho người sử dụng.

Trong ví dụ trên, báo cáo 'Thống kê doanh thu' có 4 tham số được hiển thị trên form để người dùng chọn điều kiện: Viễn thông tỉnh, năm, quý, tháng.

3.2.6. Thiết kế công cụ quản trị hệ thống

Chức năng quản trị hệ thống cho phép người quản trị hệ thống định nghĩa các quyền và phân quyền cho người dùng:

- Cập nhật nhóm người dùng.
- Cập nhật nhân viên.
- Cập nhật người dùng.
- Định nghĩa form nhập dữ liệu từ điển: Người dùng có thể tự định nghĩa form nhập cho một bảng dữ liệu nào đó (bảng chiều hoặc bảng sự kiện).
- Form chung nhập dữ liệu từ điển: Từ định nghĩa form nhập dữ liệu cho 1 bảng nào đó, chương trình tự động xây dựng form nhập dữ liệu cho bảng đó.
- Cập nhật menu: Định nghĩa các hệ thống menu khác nhau dựa trên các chức năng đã được định nghĩa.
- Cập nhật báo cáo: Mô tả phần báo cáo ở trên.
- Cập nhật hàm.
- Định nghĩa các bảng giới hạn thao tác.
- Quản lý quyền, mỗi quyền bao gồm 4 thành phần: menu, báo cáo, hàm, giá trị

bảng được phép thao tác.

- Gán quyền cho người dùng: Có thể gán 1 hoặc nhiều role cho một nhóm người dùng. Một người dùng có thể có các quyền ứng với 1 role nào đó, ngoài ra còn có thêm các quyền riêng lẻ.
- Gán quyền cho nhóm người dùng.

3.3. Cài đặt, triển khai

3.3.1. Cài đặt hệ thống

Theo hình 53, hệ thống kho dữ liệu thử nghiệm sẽ bao gồm các thành phần sau: Máy chủ lưu kho dữ liệu, máy chủ ETL, máy chủ báo cáo+ webserver, máy trạm.

- Máy chủ lưu kho dữ liệu: Có bộ nhớ tối thiểu 2GB, dung lượng ổ cứng tối thiểu là 80GB (Có thể lưu trữ được dữ liệu của vài tháng). Được cài đặt Oracle 11G. Trên máy chủ này có 3 user chính: OWBSYS – Lưu trữ cấu hình, quản trị, thiết kế kho dữ liệu; ADMIN_WH – Lưu trữ cở sở dữ liệu hệ thống báo cáo động và quản trị hệ thống; Lưu trữ WH_CCBS – Lưu trữ dữ liệu chính kho dữ liệu.
- Máy chủ ETL: Có cài đặt Warehouse Builder 11G.
- Máy chủ báo cáo + WebServer + Proccess Server: Cài đặt dotnetfx 2.0, Crystal report server 9, IIS. Có chứa các ứng dụng: Địa chỉ web báo cáo tra cứu, Proccess Server, Địa chỉ cài đặt cho máy client (Bộ cài dùng chung cho tất cả các máy trạm: bộ cài dotnetfx 2.0, crystalreports for net 2.0 và ứng dụng).
- Máy trạm: Máy trạm có thể chạy các ứng dụng báo cáo và tra cứu bằng web hoặc winform. Để chạy được ứng dụng bằng winfom phải cài đặt dotnetfx 2.0, crystalreports for net 2.0. Người dùng cài đặt qua địa chỉ web, khi có phiên bản mới tự động tải về.

Khi cài đặt chúng ta có thể cài đặt tất cả các máy chủ kho dữ liệu, ứng dụng và ETL trên cùng 1 máy.

3.3.2. Triển khai

Sau khi đã cài đặt hệ thống và có bản thiết kế DWH ở trên, việc triển khai DWH qua các bước sau:

- Tạo cấu trúc cở sở dữ liệu cho các chiều: Sử dụng OWB để delploy các chiều đã được thiết kế ở trên.
- Tạo cấu trúc cở sở dữ liệu cho các bảng sự kiện: Sử dụng OWB để delploy các bảng sự kiện đã được thiết kế ở trên.
- Tạo cấu trúc cở sở dữ liệu cho các bảng tổng hợp, view, materialized view: Sử dụng OWB để delploy các materialized view đã được thiết kế ở trên. Lúc này cơ

sở hạ tầng của DWH đã được hoàn chỉnh.

- Tạo các hàm, thủ tục, package cho các mapping thu thập số liệu: Sử dụng OWB để delploy các mapping thu thập số liệu đã được thiết kế ở trên.
- Chạy các mapping thu thập số liệu cho các chiều: Sử dụng OWB để start các mapping thu thập số liệu các chiều.
- Chạy các mapping thu thập số liệu cho các bảng sự kiện: Sử dụng OWB để start các mapping thu thập số liệu các bảng sự kiện.
- Tạo người dùng, phân quyền cho người dùng xem báo cáo, tra cứu và cập nhật số liệu.
- Tạo báo cáo, tra cứu,…phân quyền cho người dùng báo cáo, tra cứu đó.

Chương 4: KẾT QUẢ VÀ HƯỚNG PHÁT TRIỂN

4.1. Kết quả

Sau một thời gian nghiên cứu và thực hiện các nội dung của luận văn, luận văn đạt được các kết quả sau:

- Về mặt lý thuyết: Tìm hiểu về DWH, các kiến trúc công nghệ để xây dựng DWH.
- Tìm hiểu hệ thống nguồn CCBS, các yêu cầu của một hệ thống DWH viễn thông cần có.
- Có bản thiết kế chi tiết mô hình lôgíc, mô hình vật lý, thu thập số liệu cho DWH.
- Thu thập và chuyển đổi số liệu từ hệ thống CCBS đến hệ thống DWH.
- Xây dựng được công cụ báo cáo, tra cứu động, quản trị hệ thống cho DWH đã xây dựng.
- Xây dựng được một số báo cáo phân tích.

4.2. Báo cáo phân tích

Hiện tại luận văn đã thiết kế một số báo cáo, tra cứu phân tích cơ bản, việc thiết kế các báo cáo, tra cứu mới người dùng có thể thực hiện dễ dàng bằng công cụ tạo báo cáo và tra cứu động mà không phải dịch lại code.

- Báo cáo phân tích bán hàng: Thống kê phát triển thuê bao theo Viễn thông tỉnh, dịch vụ viễn thông, thời gian.
- Báo cáo thuê bao phát triển trong chương trình khuyến mãi: Thống kê phát triển thuê bao trong chương trình khuyến mãi theo Viễn thông tỉnh, dịch vụ viễn thông.

- Báo cáo tổng hợp doanh thu theo thời gian: Thống kê doanh thu theo Viễn thông tỉnh, thời gian (Năm, quý, tháng).
- Báo cáo tổng hợp thu tiền theo thời gian: Thống kê thu tiền theo Viễn thông tỉnh, thời gian (Năm, quý, tháng).
- Báo cáo tổng hợp tiền nợ theo thời gian: Thống kê tiền nợ theo Viễn thông tỉnh, thời gian (Năm, quý, tháng).
- Thống kê danh sách khách hàng có doanh thu cao nhất.
- Thống kê danh sách khách hàng có doanh cao.
- Báo cáo phân tích khiếu nại khách hàng: Thông kê khiếu nại của khách hàng theo tháng, dịch vụ viễn thông.
- Thống kê báo hỏng: Thông kê báo hỏng của khách hàng theo tháng, dịch vụ viễn thông.
- Thống kê sử dụng dịch vụ: Thông kê sử dụng dịch vụ của khách hàng theo Viễn thông tỉnh, dịch vụ viễn thông, tháng.
- Thống kê cuộc gọi theo giờ gọi: Thống kê cuộc gọi theo viễn thông tỉnh, thời gian (Năm, quý, tháng).
- Thống kê cuộc gọi theo mã vùng: Thống kê cuộc gọi theo viễn thông tỉnh, thời gian (Năm, quý, tháng).
- Thống kê cuộc gọi theo dịch vụ: Thống kê cuộc gọi theo viễn thông tỉnh, thời gian (Năm, quý, tháng).

4.3. Hướng phát triển của đề tài

Hiện tại kho dữ liệu đã thiết kế hoàn chỉnh cho doanh nghiệp viễn thông nhưng dữ liệu nguồn chủ yếu lấy từ hệ thống tính cước chăm sóc khách hàng của một Viễn thông tỉnh thành phố. Dữ liệu khách hàng sử dụng dịch vụ viễn thông của VNPT đang chủ yếu tập trung ở các tỉnh, tập đoàn Bưu chính Viễn thông đang có kế hoạch quản dữ liệu tập trung cho tất cả 64 tỉnh thành. Phần khai thác kho dữ liệu chủ yếu mới là các báo cáo và tra cứu. Từ nhu cầu thực tế của các viễn thông tỉnh và tập đoàn cũng như hoàn thiệt thiết kế kho dữ liệu, một số hướng tiếp tục của đề tài sẽ là:

- Tiếp tục xác định thêm các yêu cầu tại các Viễn thông tỉnh và trên tập đoàn để thiết kế kho dữ liệu hoàn chỉnh hơn.
- Xây dựng công cụ chuyển đổi từ nhiều nguồn khác nhau. Bổ sung dữ liệu quản lý mạng như trạng thái mạng, lưu lượng mạng,…
- Thiết kế thu thập số liệu khi có dự án quản lý khách hàng tập trung của VNPT.
- Sử dụng công cụ khai thác phân tích dữ liệu trực tuyến OLAP hỗ trợ quá trình ra quyết định.

- Xây dựng các ứng dụng khai phá dữ liệu chẳng hạn như phát hiện gian lận, quản lý chăm sóc khách hàng, phát hiện và cô lập lỗi trên hệ thống mạng viễn thông.
- Xây dựng một số phần dưới dạng kho dữ liệu thời gian thực.

KẾT LUẬN VÀ KIẾN NGHỊ

Sau một thời gian nghiên cứu và thực hiện các nội dung của luận văn, tác giả đã xây dựng được một DWH ban đầu tương đối hoàn chỉnh cho Viễn thông tỉnh lấy từ hệ thống nguồn CCBS. Hiện tại hệ thống thử nghiệm đã chạy thử nghiệm tại Viễn Long An và trong thời gian tới tiếp tục thử nghiệm tại một số Viễn thông tỉnh khác. Hệ thống tiếp tục được hoàn thiện, mở rộng để có thể áp dụng chính thức cho các Viễn thông tỉnh cũng như VNPT. Hệ thống DWH được xây dựng chung cho lĩnh vực viễn thông nên không chỉ áp dụng VNPT mà còn có thể áp dụng cho các doanh nghiệp viễn thông khác.

TÀI LIỆU THAM KHẢO

Tài liệu tiếng Việt

1. Đoàn Văn Ban (1997), *Phương pháp thiết kế và khai thác kho dữ liệu,* đề tài nghiên cứu cấp Trung tâm KHTN & CNQG

2. Hà Thái Bảo, Trung tâm Công nghệ Thông tin, Học viện Bưu chính Viễn thông (2002), *Áp dụng các công nghệ mới để xây dựng hệ thống tính cước và chăm sóc khách hàng theo mô hình nhiều lớp (Multi-layer)*, Hà Nội.

3. Nguyễn Thanh Bình, Đại học Huế, *Kho dữ liệu và Hệ hỗ trợ quyết định*, Huế.

4. Công ty Cổ phần Tin học Ứng dụng và Viễn thông (2008), *Tài liệu thiết kế hệ thống tính cước và chăm sóc khách hàng*, Hà Nội.

5. Huỳnh Đức Nghĩa, Trung tâm Công nghệ Thông tin, Học viện Bưu chính Viễn thông (2004), *Nghiên cứu công nghệ phân tán trên nền .NET Framework áp dụng cho việc phát triển và tích hợp các hệ thống phần mềm*, Hà Nội.

6. Tập đoàn Bưu chính Viễn thông Việt Nam VNPT (2006), *Tài liệu quy trình nghiệp vụ*, Hà Nội.

Tài liệu tiếng Anh.

7. Artech House, Boston London (1997), *Data Warehousing and Data Mining for Telecommunications*, London.

8. IBM (2003), *Telecommunications Data Warehouse*, USA..

9. J.D. Meier, Alex Mackman, Michael Dunner, and Srinath Vasireddy, Microsoft Corporation (2002), *.NET Remoting Security.*

10. Oracle (2007), *Data Warehousing Guide 11g Release 1*, USA.

11. VINCENT RAINARDI (2008), *Building a Data Warehouse With Examples in SQL Server*, USA.

12. Ralph Kimball Margy Ross (2002), *The Data WarehouseToolkit Second Edition The Complete Guide to Dimensional Modeling*, New York.

13. Stephen Toub (2002), *"Secure Your .NET Remoting Traffic by Writing an Asymmetric Encryption Channel Sink"*, MSDN Magazine.

14. Oracle (2007), *Warehou Builder Guide 11g Release 1*, USA.

15. WWW.ORACLE.COM

CÁC THUẬT NGỮ VÀ VIẾT TẮT

- CCBS (Customer Care and Billing System): Hệ thống Tính cước và Chăm sóc khách hàng.
- CSDL: Cơ sở dữ liệu.
- DM (Data mart): Kho dữ liệu chủ đề.
- DWH (Data Warehouse): Kho dữ liệu.
- ETL (Extract Transform Load): Trích xuất chuyển đổi nạp dữ liệu.
- OLTP (Online Transaction Processing): Xử lý giao dịch trực tuyến.
- OLAP (Online Analytical Processing): Xử lý phân tích trực tuyến.
- OWB (Oracle Warehouse Builder): Công cụ xây dựng kho dữ liệu của Oracle.
- VNPT (Vietnam Posts and Telecommunications Group): Tập đoàn Bưu chính Viễn thông Việt nam.