

HỆ THỐNG PHÂN TÍCH KINH DOANH DI ĐỘNG

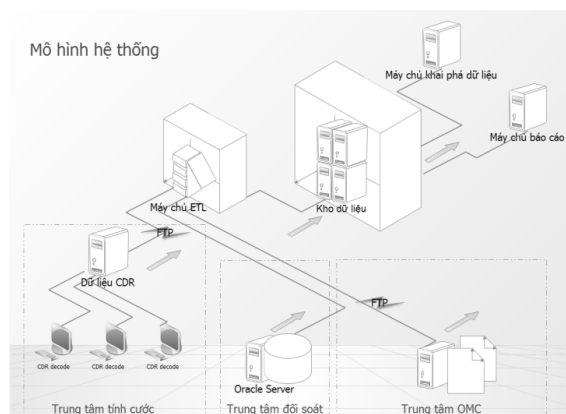
ThS. Nguyễn Anh Nguyễn, ThS. Phạm Đức Tú

Phòng NCPT Ứng dụng Viễn thông

Tóm tắt: VNPT trong những năm qua đang lưu trữ một khối lượng dữ liệu khổng lồ, bao gồm các dữ liệu như thông tin về khách hàng, chi tiết cuộc gọi, các dịch vụ thuê bao, thông tin cảnh báo tình trạng hệ thống mạng viễn thông, dữ liệu về hệ thống cấp viễn thông v.v... Lượng dữ liệu khổng lồ này nếu được khai thác đúng cách sẽ là một lợi thế cho các doanh nghiệp viễn thông trong việc tung ra các dịch vụ chăm sóc khách hàng hay các ứng dụng mang tính thực tiễn cao, đồng nghĩa với việc tăng cơ hội cạnh tranh đối với các doanh nghiệp khác. Từ đó yêu cầu đặt ra việc xây dựng kho dữ liệu từ những nguồn dữ liệu trên phục vụ cho việc phân tích kinh doanh tại các đơn vị viễn thông. Tuy nhiên việc xây dựng một hệ thống như thế vấp phải một số hạn chế về mặt kỹ thuật, đặc biệt là khi kích thước cũng như độ phức tạp của môi trường thông tin tăng lên. Xét về mặt lưu trữ, dữ liệu trong các hệ thống tác nghiệp được trữ phân tán ở nhiều dạng không tương thích với nhau, thậm chí còn ở những dạng phi cấu trúc. Nhiều hệ CSDL đã được xây dựng không tương thích với nhau và không tương thích với những hệ thống tin mới được xây dựng. Từ đó yêu cầu đặt ra là làm thế nào để trích rút những thông tin có ích từ những hệ thống này, tích hợp thành các thông tin chung để có thể đưa vào kho dữ liệu để khai thác. Xuất phát từ khó khăn đó, tập đoàn đã chỉ đạo thực hiện giải pháp tiếp nhận trực tiếp số liệu thô về cước di động của Vinaphone và Mobifone, phân tích xử lý và cung cấp ngay các báo cáo về các dịch vụ như Voice, SMS, Data, Vas của di động hàng ngày cho các cấp lãnh đạo nhằm hỗ trợ đưa ra quyết định kịp thời trong kinh doanh. Bài báo này nói về một số giải pháp trong việc phân tích, xử lý và cung cấp các báo cáo kinh doanh của VNPT về lĩnh vực di động do Viện công nghệ thông tin và truyền thông CDIT thực hiện.

1. MÔ HÌNH KIẾN TRÚC

Dữ liệu từ các hệ thống của Vinaphone gồm dữ liệu chi tiết cước, dữ liệu lưu lượng tổng đài, dữ liệu trừ tiền thuê bao trả trước và dữ liệu nạp thẻ. Dữ liệu này được lưu trữ trong nhiều hệ thống với nhiều khuôn dạng khác nhau. Mô hình hệ thống được mô tả như hình 3.1. Dữ liệu tiêu thụ thuê bao trả trước lưu trong cơ sở dữ liệu Oracle bao gồm:



- **Dữ liệu tiêu thụ cuộc gọi và SMS (DATA_CALL_HISTORY):** Dữ liệu tiêu thụ trên các loại tài khoản của khách

hàng, cho biết thuê bao phát sinh, loại cuộc gọi hay SMS, số tiền ban đầu và số tiền bị trừ trong mỗi loại tài khoản, thông tin về Location phát sinh trong trường hợp loại thuê bao cuộc gọi. Mỗi ngày hệ thống yêu cầu xử lý khoảng 200.000.000 bản ghi với dung lượng 80 GB dữ liệu.

- **Dữ liệu tiêu thụ Data (3G+GPRS) và đăng ký gói cước cập nhật theo ngày.** Các dữ liệu này chứa thông tin về chi tiết trừ tiền khi các thuê bao trả trước truy cập internet hay khi các thuê bao đăng ký các gói dịch vụ của Vinaphone. Các thông tin này bao gồm: Số thuê bao đăng ký (truy cập), loại gói (các gói cước hoặc loại truy cập internet), thông tin chi tiết gói (các loại gói con như M10, M35, M50...), lưu lượng truy cập data bao dữ liệu tiêu thụ trên 3G và GPRS. Mỗi ngày hệ thống yêu cầu xử lý khoảng 80.000.000 bản ghi với dung lượng 20 GB dữ liệu.

- **Dữ liệu CDR:** Lưu trữ dưới dạng file FoxPro, thông tin lưu trữ bao gồm chi tiết

cước cuộc gọi và SMS của cả thuê bao trả sau và trả trước. Mỗi ngày hệ thống yêu cầu xử lý khoảng 400.000.000 bản ghi với dung lượng 60 GB dữ liệu.

2. CÁC KỸ THUẬT SỬ DỤNG

Với nguồn dữ liệu đầu vào hàng tỉ bản ghi với hàng trăm GB, và để đưa ra được các thông tin cho các chỉ tiêu với các nhóm dịch vụ VOICE, SMS, DATA, VAS phục vụ cho điều hành sản xuất kinh doanh tập đoàn trong giới hạn thời gian nhóm triển khai đã áp dụng một số kỹ thuật giúp tối ưu và đảm bảo yêu cầu bài toán.

2.1. Kỹ thuật ETL

Kỹ thuật nội suy dữ liệu location, cho phép xác định lac, cell với độ chính xác lên tới 90% cho các dữ liệu không có thông tin về location.

Kỹ thuật chia để trị với việc tải dữ liệu, cho phép tối ưu hóa thời gian tải dữ liệu thông qua sử dụng đa luồng.

2.2. Kỹ thuật Partitioning và materialized views

Khi dữ liệu di động được lưu trữ ngày một lớn thì việc áp dụng kỹ thuật Partitioning là rất quan trọng trong tối ưu việc lưu trữ và

tăng tốc độ xử lý của hệ thống. Với kỹ thuật này thì đảm bảo hệ thống có thể lưu trữ và xử lý dữ liệu trên hàng tỷ bản ghi với tốc độ đảm bảo yêu cầu.

Áp dụng Materialized views trong kỹ thuật xử lý báo cáo, giúp dữ liệu tổng hợp một cách tự động ngay khi có sự thay đổi ở dữ liệu nguồn; tối ưu hóa thời gian tính toán và đảm bảo sự đồng bộ giữa dữ liệu nguồn và dữ liệu tổng hợp.

3. KẾT LUẬN

Hệ thống phân tích kinh doanh đã được triển khai thành công đáp ứng cho nhu cầu phân tích kinh doanh không chỉ cho lãnh đạo VNPT mà còn hỗ trợ cho 64 đơn vị VNPT tỉnh thành đưa ra những chiến lược hợp lý cho sản xuất kinh doanh.

Việc triển khai thành công hệ thống này không chỉ là kết quả thành công về mặt triển khai ứng dụng mà còn là kết quả thành công trong việc quản lý các ứng dụng dữ liệu lớn.

Trong tương lai nhóm triển khai đề xuất phát triển tiếp việc khai phá dữ liệu và hoàn chỉnh quy trình xây dựng hệ thống tiến đến xây dựng các hệ thống không chỉ trong lĩnh vực viễn thông mà trong nhiều lĩnh vực khác nhau.

Thông tin tác giả:



Phạm Đức Tú

Năm sinh : 1977

Lý lịch khoa học:

- 1995-1999 : Khoa CNTT, đại học Khoa học tự nhiên, đại học quốc gia Hà nội

- 2011-2013 : Thạc sĩ, khoa học máy tính, Học viện CNBCVT

Hướng NC đang theo đuổi : Công nghệ trong lĩnh vực xử lý BigData

Email : tupd@ptit.edu.vn - phamductu@cdit.com.vn

Nguyễn Anh Nguyên

Năm sinh : 05/01/1984

Lý lịch khoa học:

- 2002-2007 : Khoa CNTT, đại học Bách khoa Hà nội

- 2010-2012 : Thạc sĩ, khoa học máy tính, Học viện CNBCVT

Hướng NC đang theo đuổi : Nghiên cứu công nghệ phần mềm liên quan tới các thiết bị di động

Email : nguyenna@ptit.edu.vn