An abstract graphic featuring a central point from which several wide, colorful rays (green, orange, blue, grey) emanate outwards. The background is a light grey grid. Faint binary code (0s and 1s) and a small line graph are visible in the upper right area.

HỆ THỐNG THÔNG TIN PHỤC VỤ TRÍ TUỆ KINH DOANH

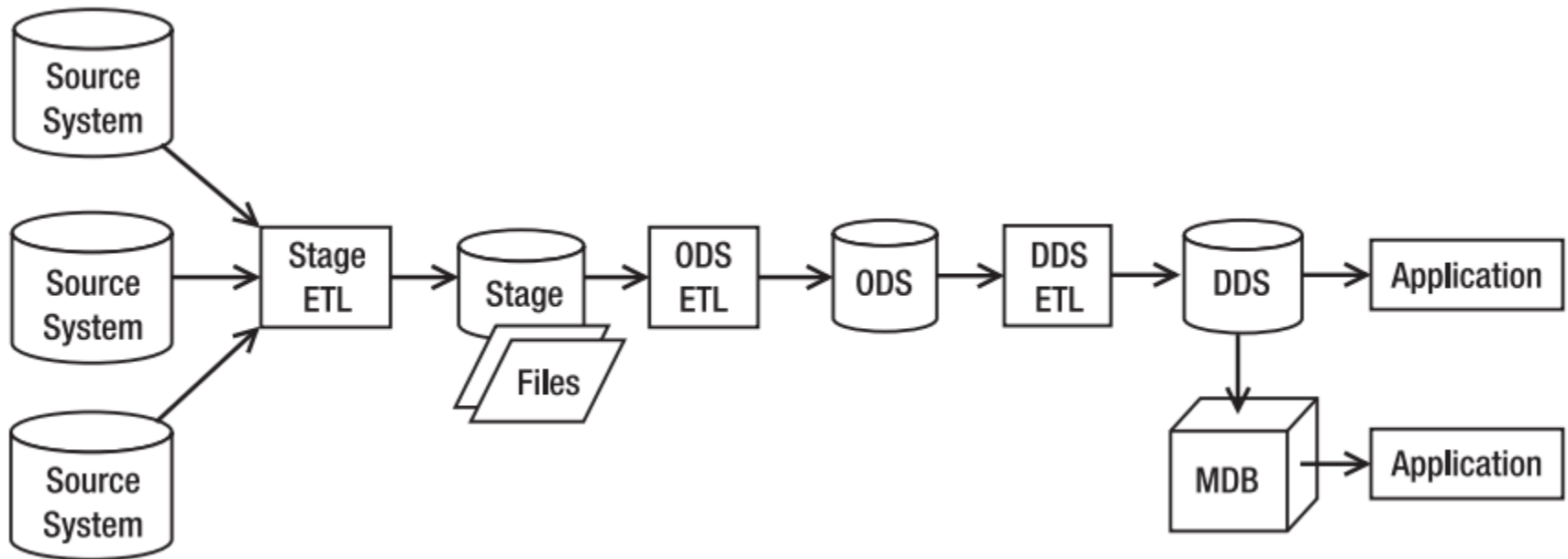
MÔ HÌNH HOÁ DỮ LIỆU – Case study(tt)

TB: Hồ Thị Hoàng Vy

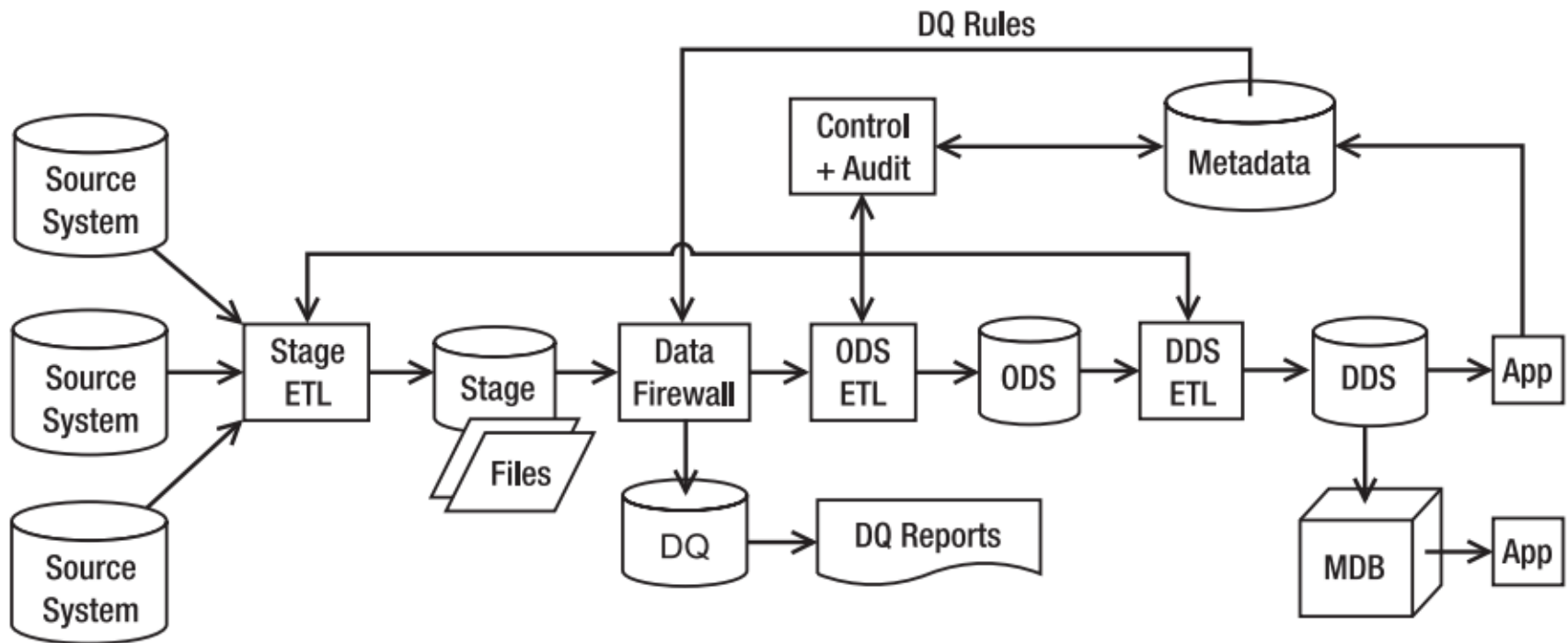
2013, Jan 27

➔ Data flow architecture

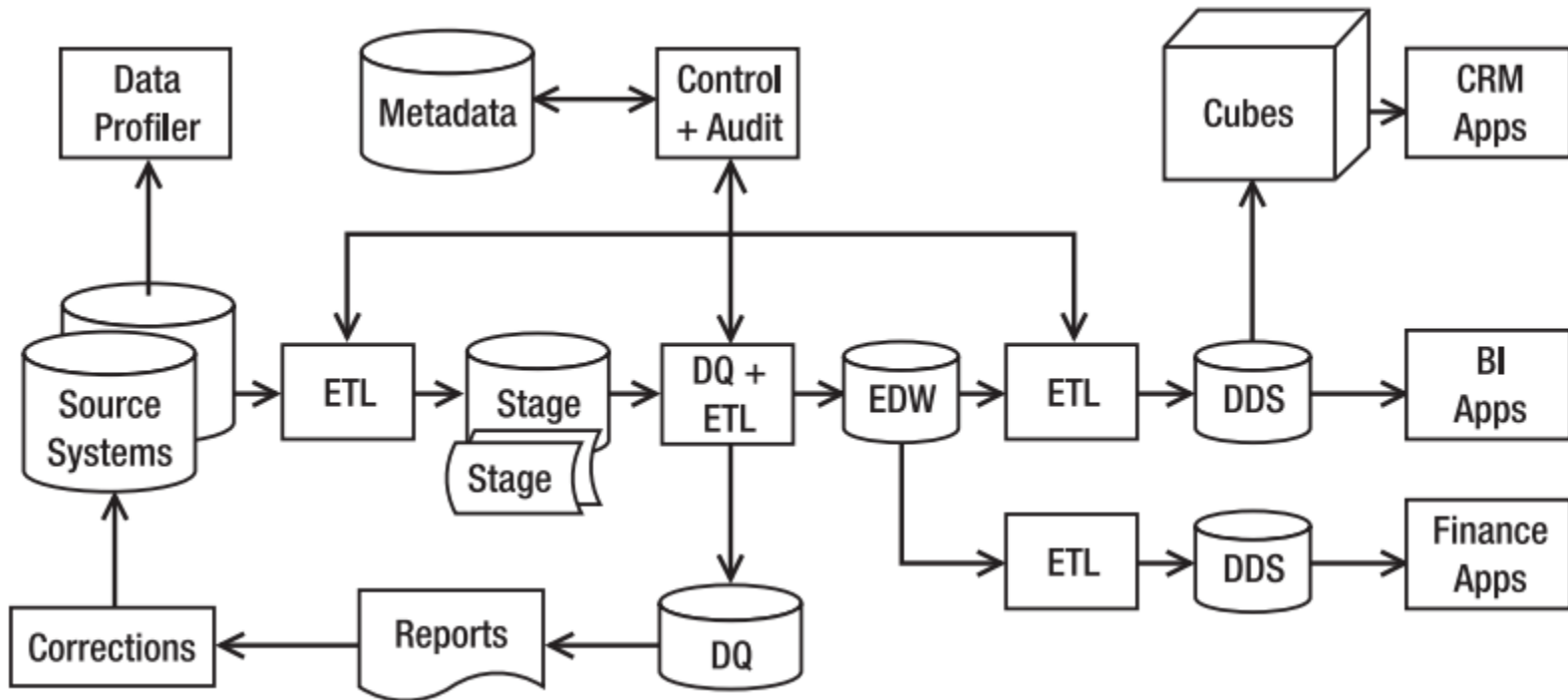
- VD data flow architecture



➔ Data flow architecture

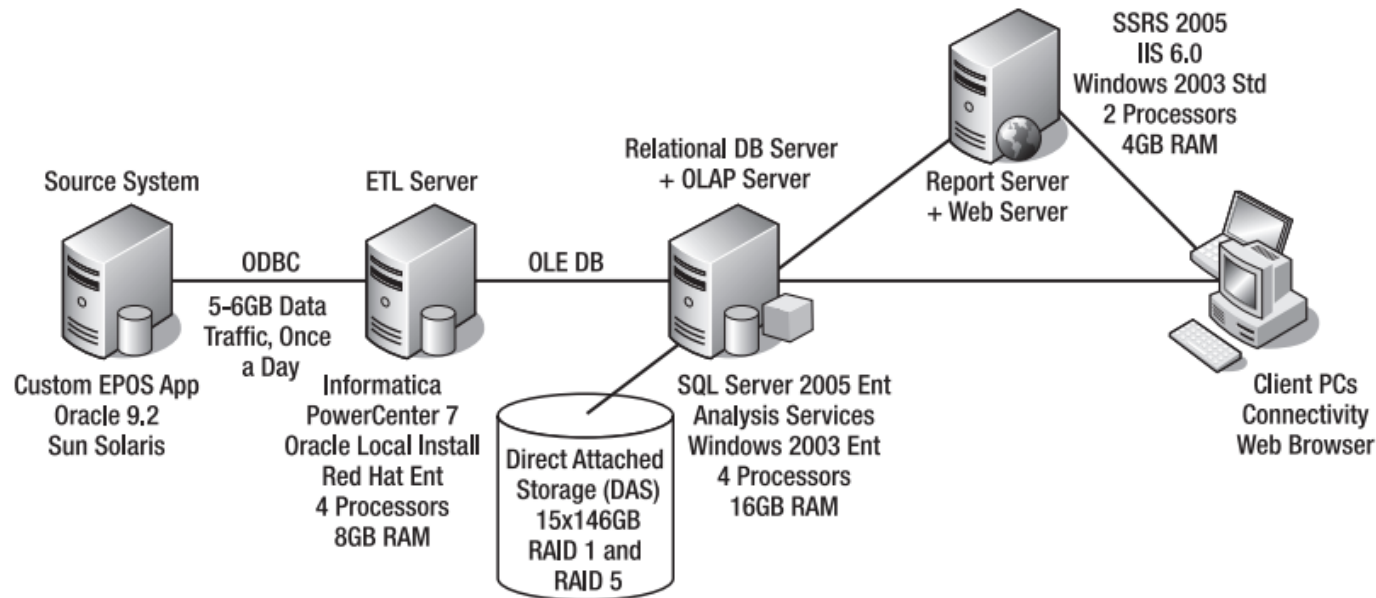


Data flow architecture



➔ System architecture

- Sau khi chọn được kiến trúc luồng dữ liệu → thiết kế kiến trúc hệ thống:
 - Sự sắp xếp vật lý, các kết nối giữa máy chủ, mạng, phần mềm, hệ thống lưu trữ, máy trạm



➔ Thiết kế KDL

- Nghiên cứu tính khả thi

- Khảo sát nhu cầu cấp cao
- Đánh giá chất lượng dữ liệu
- Viết các business case
- Lập kế hoạch dự án
-

➔ Hiểu rõ nhu cầu, đề xuất giải pháp, kiến trúc dữ liệu và hệ thống, hạ tầng và chất lượng dữ liệu để đánh giá thời gian và chi phí thực hiện

➔ Thiết kế KDL

- Tìm hiểu nhu cầu:
 - Chi tiết các quy trình, nghiệp vụ, dữ liệu, các vấn đề
 - Thu thập các kinh nghiệm người dung, thảo luận ngữ nghĩa dữ liệu, giao diện..
 - Liệt kê các nhu cầu chức năng & phi chức năng, hiệu suất và bảo mật

➔ Thiết kế KDL

- **Kiến trúc:**

- Xác định kiến trúc hệ thống đang dung
 - Đặc tả chi tiết: server dữ liệu, loại hình mạng, giải pháp lưu trữ,
 - Quyết định kiến trúc dữ liệu sẽ dung
- ➔ Trả lời các câu hỏi:
- Không gian lưu trữ
 - Khả năng xử lý

➔ Thiết kế KDL

- **Thiết kế**

- Kho chứa dữ liệu (data store)
- Hệ thống ETL
- Các ứng dụng Front-end
- Hệ thống đảm bảo chất lượng dữ liệu (data quality system) và meta data

- **Cài đặt**

- **Kiểm chứng**

- **Triển khai**

➔ Thiết kế KDL

- **Vận hành**

- Quản trị kho và hỗ trợ người dung
 - Giải quyết lỗi và vấn đề khi sử dụng hệ thống
 - Quản trị người dung mới
 - Nhu cầu phát sinh: thêm dữ liệu mới, thêm yêu cầu, , tạo report mới, cube mới, thêm tính năng trong ứng dụng front-end...

- **Thiết lập hạ tầng**

- Mua/cài đặt/cấu hình/kiểm chứng phần cứng, phần mềm

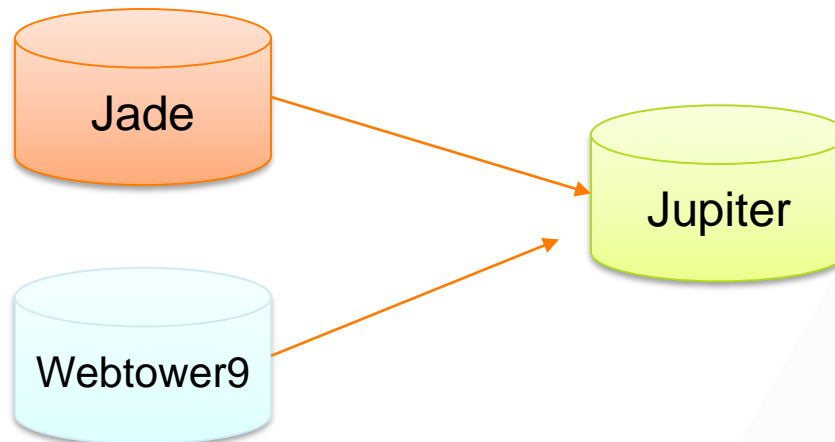
- **Quản lý dự án**

➡ VẬN DỤNG

- Phân tích case study
 - Đề án lý thuyết

➔ Hiện trạng

- Công ty Amadeus Entertainment có 3 nguồn dữ liệu:
 - Jade (DB2) – Product Sale Offline
 - Jupiter (SQL server) – Inventory & products
 - Webtower9 (Oracle) – Product sales (online)



Hiện trạng

- Công ty là một nhà bán lẻ chủ yếu: films, audio books, music.
 - Có tám cửa hàng online vận hành ở Mỹ, Đức, pháp, Anh, Tây Ban Nha, Úc, Nhật và Ấn Độ.
 - Có 96 cửa hàng offline cũng trên những quốc gia này.
 - Khách hàng có thể mua hàng cá nhân với các sản phẩm: bài hát, audio book, phim, hoặc đăng ký gói cho phép download một số sản phẩm trong 1 khoảng thời gian nhất định
 - Công ty có nhiều kênh phân phối: TV, internet, mobile phone...
 - Có nhiều phương thức chi trả: trả hàng năm, trả trước hàng tháng...

Nhu cầu

1. Người dung nghiệp vụ cần phân tích bán hàng (product sales). Cần biết được doanh số, chi phí, lợi nhuận theo một loại tiền tệ (\$)
2. Người dung nghiệp vụ cần phân tích “bán hàng qua hình thức đăng ký” (subscription sales). Cần biết được doanh số, chi phí, lợi nhuận được đánh giá hằng ngày trong khoảng 1 tháng
3. Cấp cửa hàng, có thể xem dữ liệu hằng ngày trong một vài tuần trước đó ở mức tổng quát hoặc chi tiết về doanh số, chi phí, lợi nhuận để hiểu nguyên nhân của việc doanh số thấp, lợi nhuận thấp, sản phẩm nào, khách hàng nào... là nguyên nhân của vấn đề

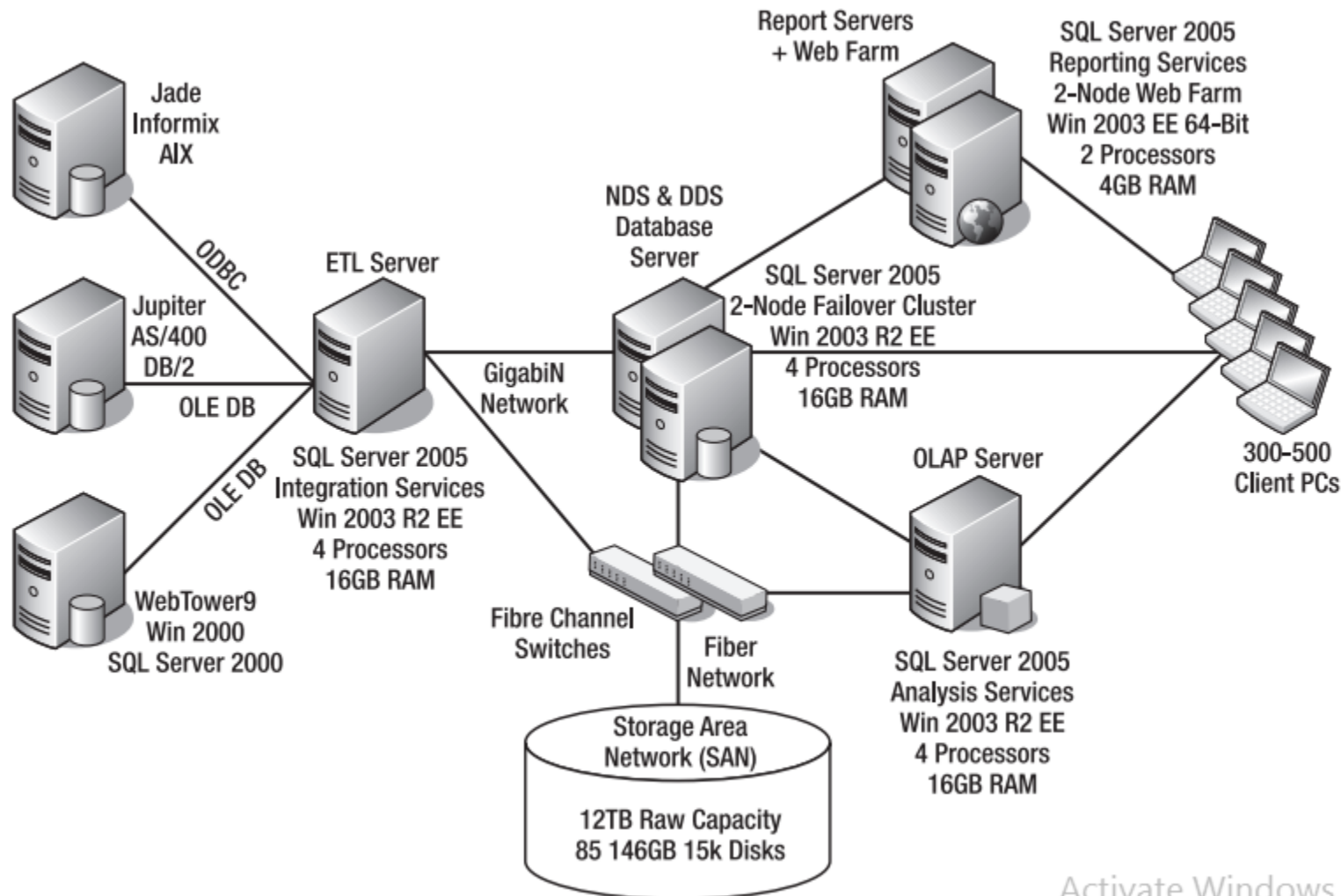
Yêu cầu

1. Cấp quản lý toàn cục có thể hiểu được xu hướng chung tổng thể hoặc từng quốc gia. Liệu một cửa hàng hay một quốc gia đang có vấn đề với một sản phẩm nào để có thể tương tác sớm với các cửa hàng nhất có thể
2. Hệ thống có thể hiển thị các hình, biểu đồ cho phép in được, có thể xuất ra excel hoặc csv..
3. Phục vụ phân tích các chiến dịch chăm sóc khách hàng qua việc xem các đo lường của mỗi chiến dịch sau khi gửi cho khách hàng: số lượng thông điệp được gửi của mỗi kênh (mobile message, email, post...), số thông điệp phân phối thành công, thất bại, tỷ lệ mở, tỷ lệ bỏ qua, than phiền, đánh spam, ...

Yêu cầu

1. Phục vụ phân đoạn khách hàng cho các chiến dịch chăm sóc khách hàng dựa vào quyền thông tin, đặc trưng địa lý, thông tin khách hàng, sở thích, lịch sử mua hàng...
2. Phân tích hiệu suất nhà cung ứng (supplier performance) – trung bình chi phí, giá trị trả về, từ chối, thời gian ngừng hoạt động..

➔ Kiến trúc cài đặt



➔ Kiến trúc luồng dữ liệu

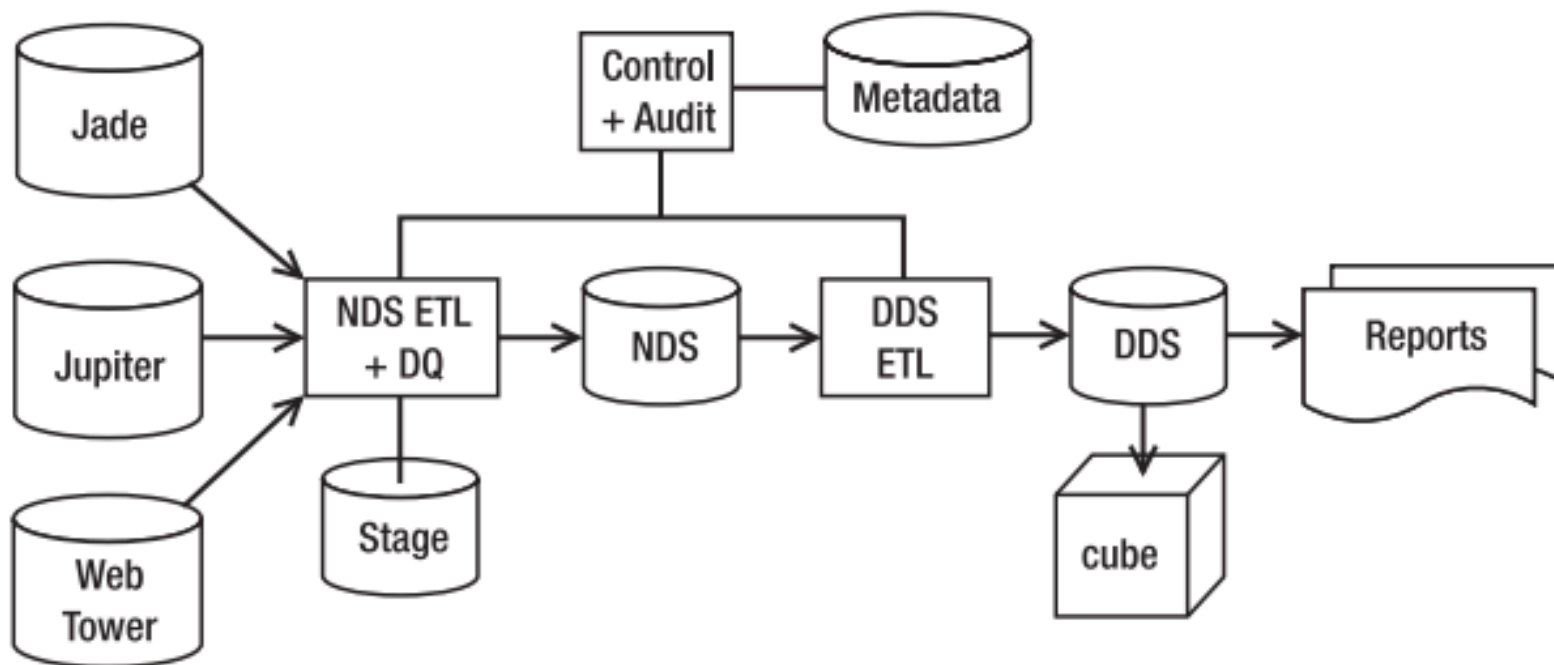


Figure 3-6. *Data flow architecture built in iteration 1*

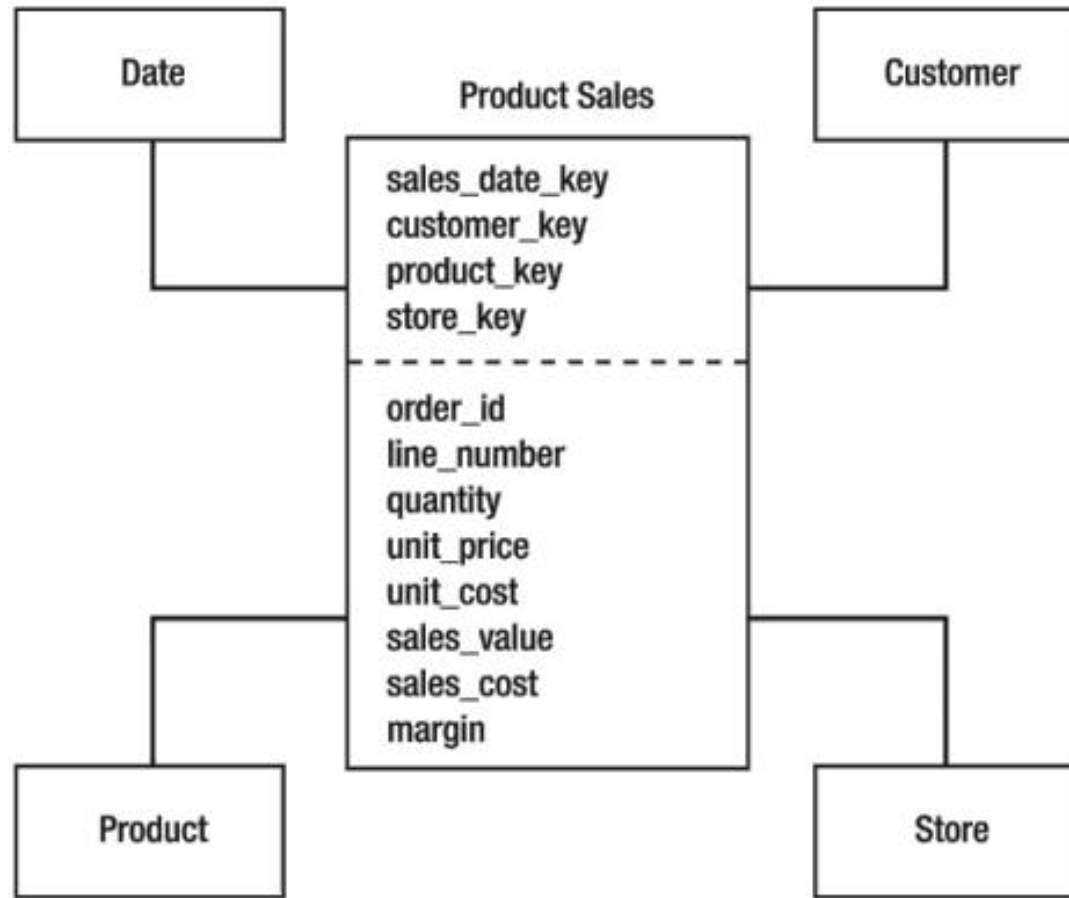
➔ Phân tích yêu cầu

- Người dung nghiệp vụ cần phân tích bán hàng (product sales). Cần biết được doanh số, chi phí, lợi nhuận theo một loại tiền tệ (\$)
 - **Sự kiện:**
 - Khi 1 khách hàng mua 1 sản phẩm
 - **Bối cảnh sự kiện:**
 - Ai: khách hàng
 - Ở đâu: cửa hàng, lãnh thổ bán hàng
 - Cái gì: sản phẩm
 - Khi nào: ngày mua hàng
 - **Đo lường** (dữ kiện): Số lượng, đơn giá, giá trị

Mô hình hoá

- Đo lường → Fact Table
- Bối cảnh → dimension table
- Mô hình sao liên kết các dimension table và fact table
- Phân cấp dữ liệu

➔ Mô hình hoá



➡ Mô hình hoá - **Fact table**

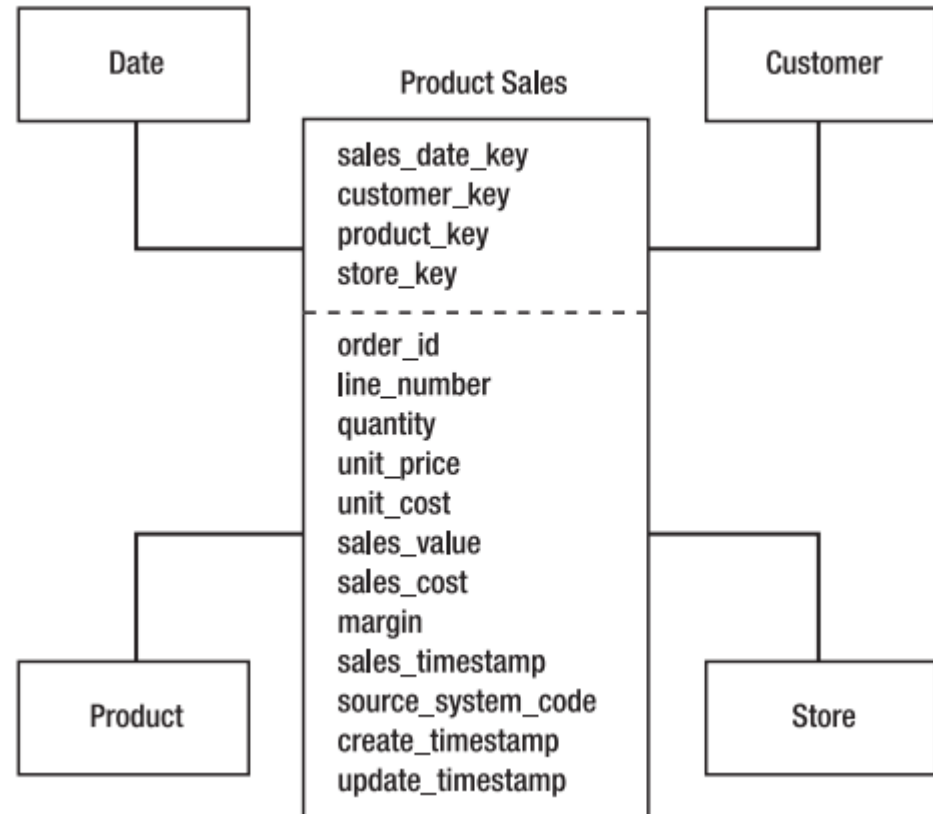
- Các giá trị có sẵn từ nguồn:
 - Quantity, unit_price, unit_cost
- Các giá trị phải tính toán:
 - Sales_value, sales_cost, margin
 - $\text{sales_value} = \text{unit_price} \times \text{quantity}$
 - $\text{sales_cost} = \text{unit_cost} \times \text{quantity}$
 - $\text{Margin} = \text{sales_value} - \text{sales_cost}$
- Cấp chi tiết dữ liệu (độ mịn)
 - Đơn vị nhỏ nhất xảy ra sự kiện:
 - Một dòng trong fact tương ứng mỗi item được bán

➡ Mô hình hoá - **Fact table**

- Nhu cầu phức tạp hơn: discounts, taxes, promotional items?
 - ➔ Xem lại nguồn dữ liệu
 - ➔ Luôn xác nhận mô hình dữ liệu với các logic nghiệp vụ cùng hệ thống nguồn để đảm bảo kho dữ liệu phản ánh đúng các điều kiện nghiệp vụ

➔ Mô hình hoá - **Fact table**

- Chiều thoái hoá
- Nhãn thời gian
- Khoá chính của fact?
 - 1 khách hàng A, mua bài hát B, tại cửa hàng C vào ngày D lúc 10am. Sau đó quay lại lúc 7pm để mua cùng bài hát lần nữa?



DATA MART

➡ Mô hình hoá - **Fact table**

- Order_ID, line_number
 - WebTower9 or Jade order ID;
 - WebTower9 or Jade order line number
- Source_system_code
 - Phân biệt dữ liệu từ nguồn nào (key của nguồn)
- Sales_timestamp
 - Thời gian khách hàng mua sản phẩm
- Create_timestamp – update_timestamp
 - Thời gian dòng dữ liệu được tạo/cập nhật trong DDS

➡ Mô hình hoá - **Thiết kế chiều**

- Các chiều liên quan sự kiện phân tích:
 - Date dimension
 - Product dimension
 - Customer dimension
 - Store dimension
- Lưu giá trị cũ của chiều
 - Xem chiều thay đổi chậm

➡ Mô hình hoá - **Thiết kế chiều**

– Date dimension

- Tại sao không lưu time dimension thay vì date dimension

➡ Mô hình hoá - **Thiết kế chiều**

- **Product dimension**

- Không yêu cầu lưu giá trị cũ
 - » SCD 1 - Ghi chồng giá trị cũ
- Chỉ có 1 nguồn là từ Jupiter
 - » chuyển product từ Jupiter → Product dimension
 - » Nếu từ nguồn bảng product được chuẩn hoá → phi chuẩn → join các bảng với nhau

➡ Mô hình hoá - **Thiết kế chiều**

- Customer dimension
 - Cần lưu trữ giá trị cũ:
 - occupation, household_income, address1, address2, address3, address4, city, state, zipcode, country, phone_number, email_address
 - SCD loại 2
 - Nếu cần lưu nhiều địa chỉ, email, điện thoại
 - Lưu nhiều thuộc tính (star schema)
 - Chuẩn hoá thành các bảng con (snowflake schema)

➡ Mô hình hoá - **Thiết kế chiều**

- **Store dimension:**

- Cần lưu giá trị cũ:

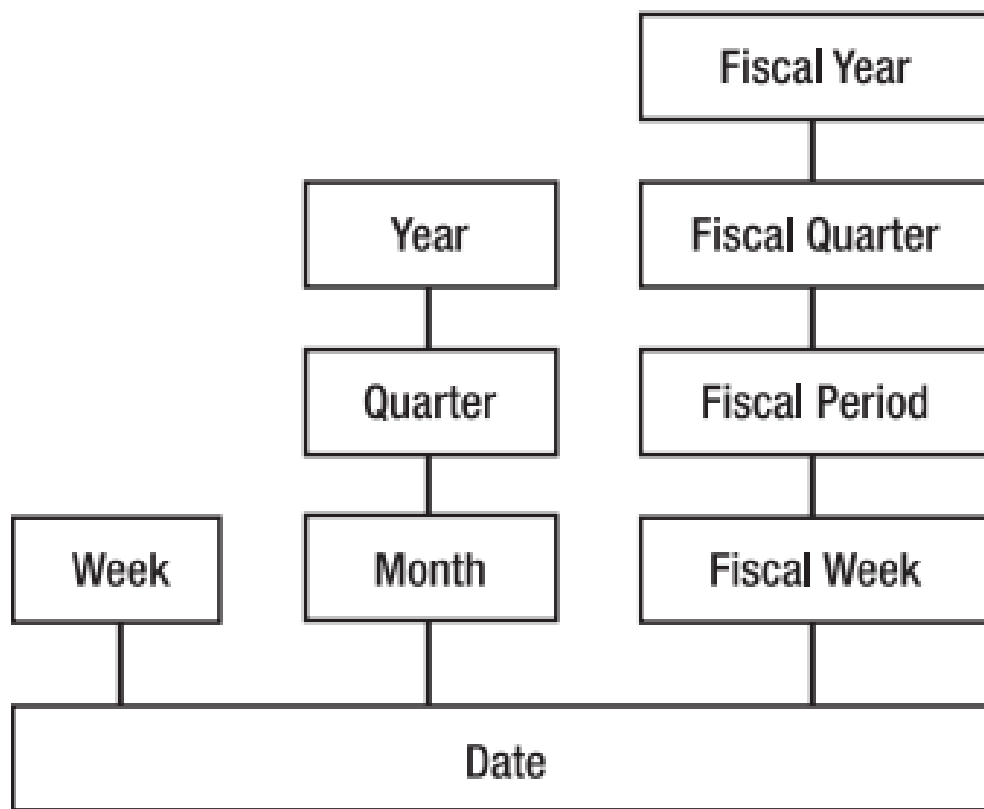
- Hiện tại có 5 region, tương lai có thể tăng thêm
 - Các cửa hàng online hiện chỉ phân bố trong 1 division, tương lai có thể phân bố vật lý như các cửa hàng offline (hiếm xảy ra)
 - SCD loại 3 → không thay đổi nhiều thuộc tính, biết chắc số lượng phiên bản thay đổi
 - Region & prior_region & prior_region_date
 - Division & prior_division & prior_division_date

➡ Mô hình hoá – Phân cấp dữ liệu

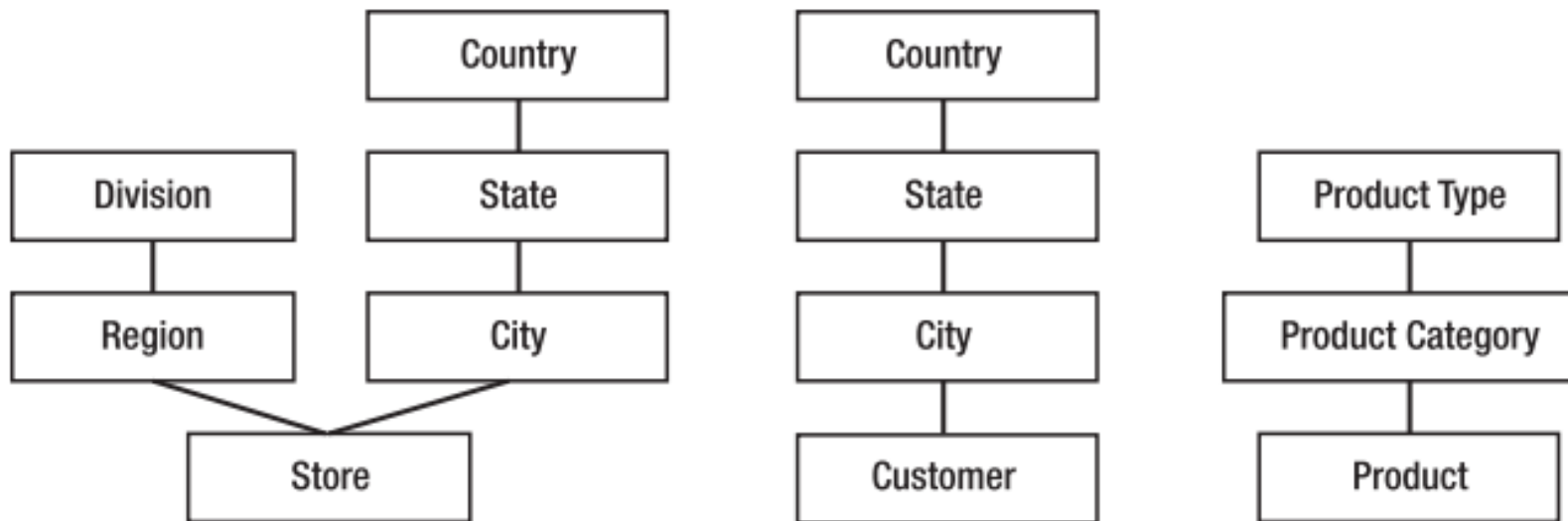
- Phục vụ nhu cầu rollup, drilldown dữ liệu
- Trong bảng chiều, đôi khi một thuộc tính là tập con của 1 thuộc tính khác
 - Giá trị của thuộc tính được gom nhóm bởi thuộc tính khác
 - Thuộc tính cấp cao: thuộc tính bị dung để gom nhóm
 - Vd: 1 năm có 4 quý, 1 quý gồm 3 tháng → cấp cao nhất là năm, tiếp theo là quý và tháng
 - Nếu biết được doanh thu bán hàng của mỗi tháng trong quý → biết được doanh số của quý đó

➔ Mô hình hoá – **Phân cấp dữ liệu**

- Chiều date



➡ Mô hình hoá – Phân cấp dữ liệu



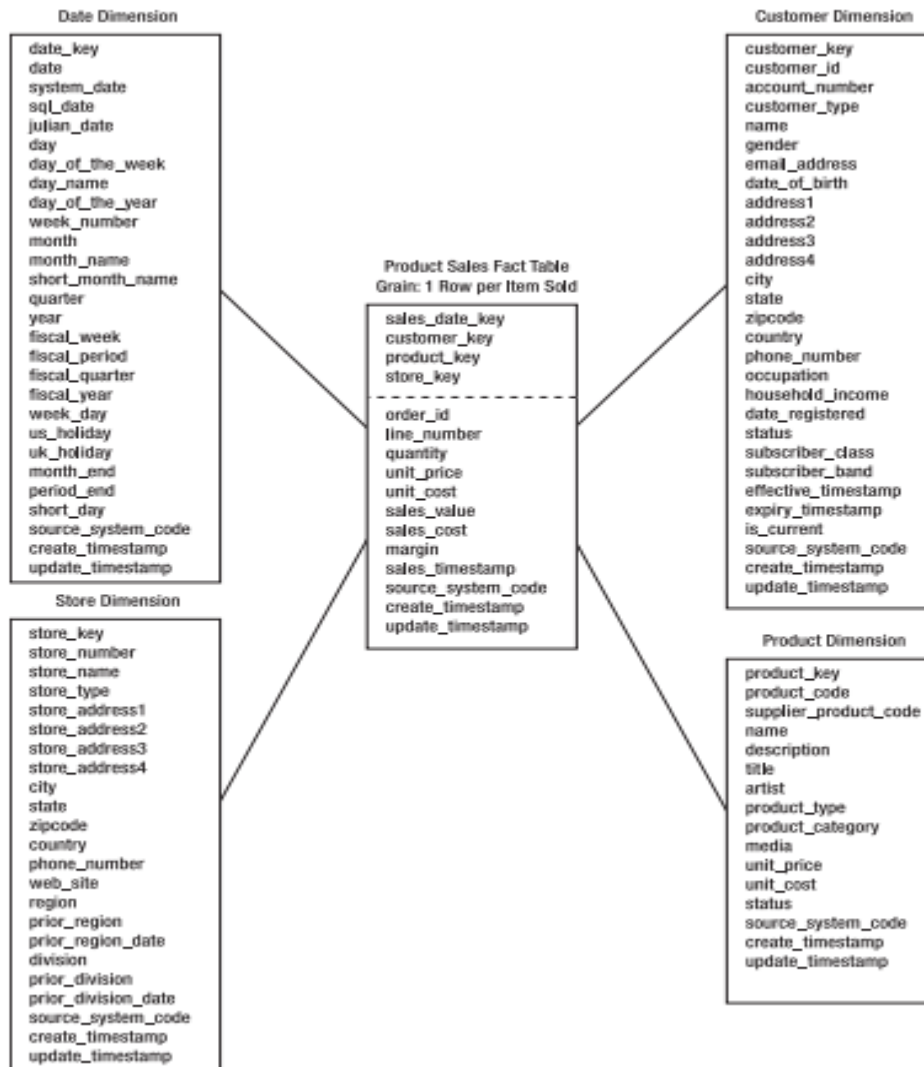
- Chiều store – customer - product

➡ Mô hình hoá

- Sau khi thiết kế xong “product sales” data mart → tiếp tục thiết kế các data mart khác theo cách tương tự:
 - Xác định fact table
 - Thiết lập mức chi tiết (độ mịn dữ liệu)
 - Xác định dimension table + SCD
 - Xây dựng lược đồ (sao, bông tuyết...)



Mô hình hoá



➔ **Ánh xạ dữ liệu nguồn**

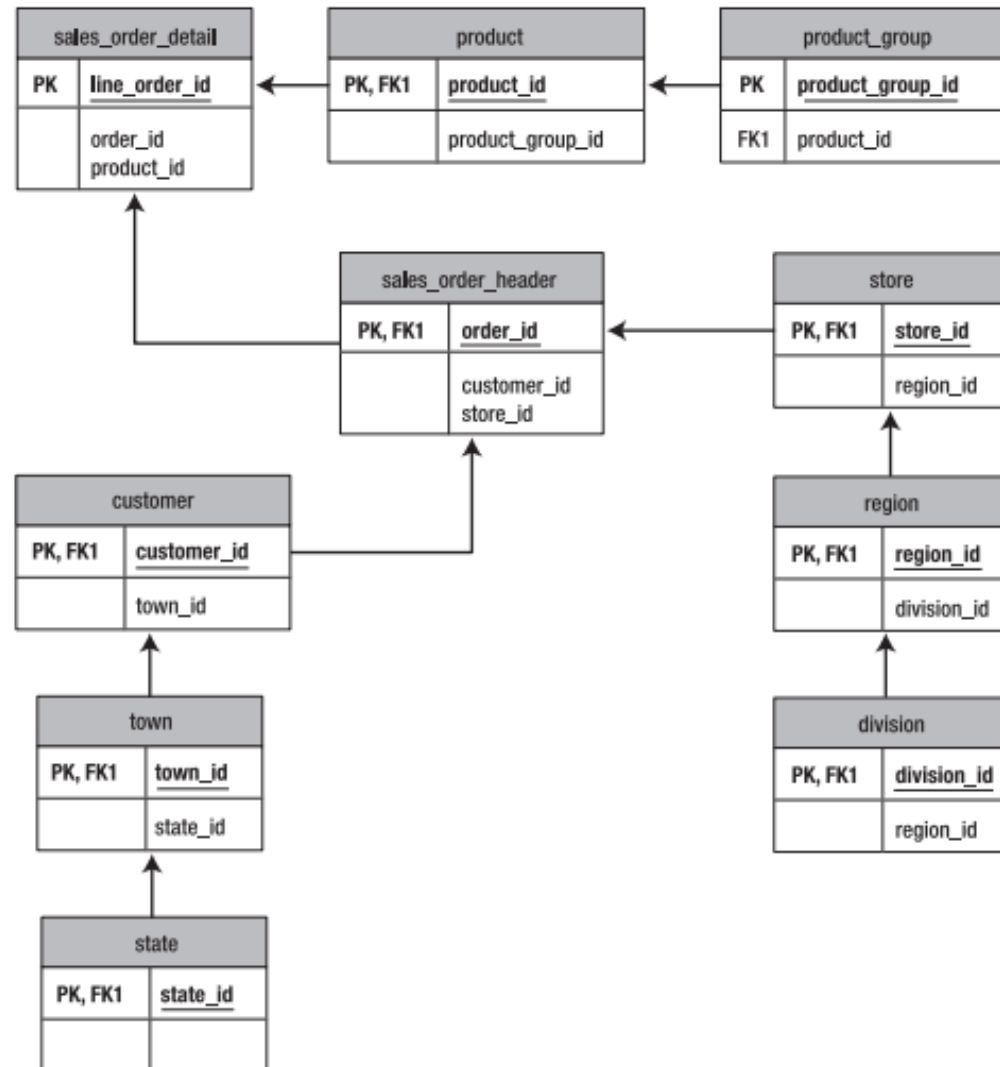
- **Sau khi thiết kế các data mart, cần ánh xạ dữ liệu DDS và nguồn**
 - Ánh xạ mỗi column trong fact table và dimension table vào dữ liệu nguồn
- **Mục đích: để xác định dữ liệu mỗi column trong DDS sẽ đến từ đâu**
 - Xác định các tính toán, chuyển đổi nếu có → giúp cho ETL
 - 1 dds column có thể đến từ nhiều table ở nguồn

➔ Thiết kế NDS

- **Liệt kê:**
 - Các bảng nguồn
 - Các bảng chuẩn hoá từ fact và dimension table
- **Kết hợp hai danh sách, phân chia thành các chủ đề khác nhau:**
 - Thiết lập mối liên hệ giữa các thực thể
 - Thiết kế khoá
 - Kết nối bảng cha và con (fact → con; dimension → cha trong NDS)
 - Chuẩn hoá dữ liệu các bảng từ 3NF trở lên

➔ Thiết kế NDS

- Ví dụ



➡ Thiết kế NDS

- **Lập bảng mô tả cho các bảng trong NDS**
 - Tên cột, kiểu dữ liệu, mô tả, nguồn, hình thức biến đổi
- Quay lại DDS, **định nghĩa lại nguồn của mỗi cột của DDS trong NDS** phục vụ cho ETL giữa NDS và DDS



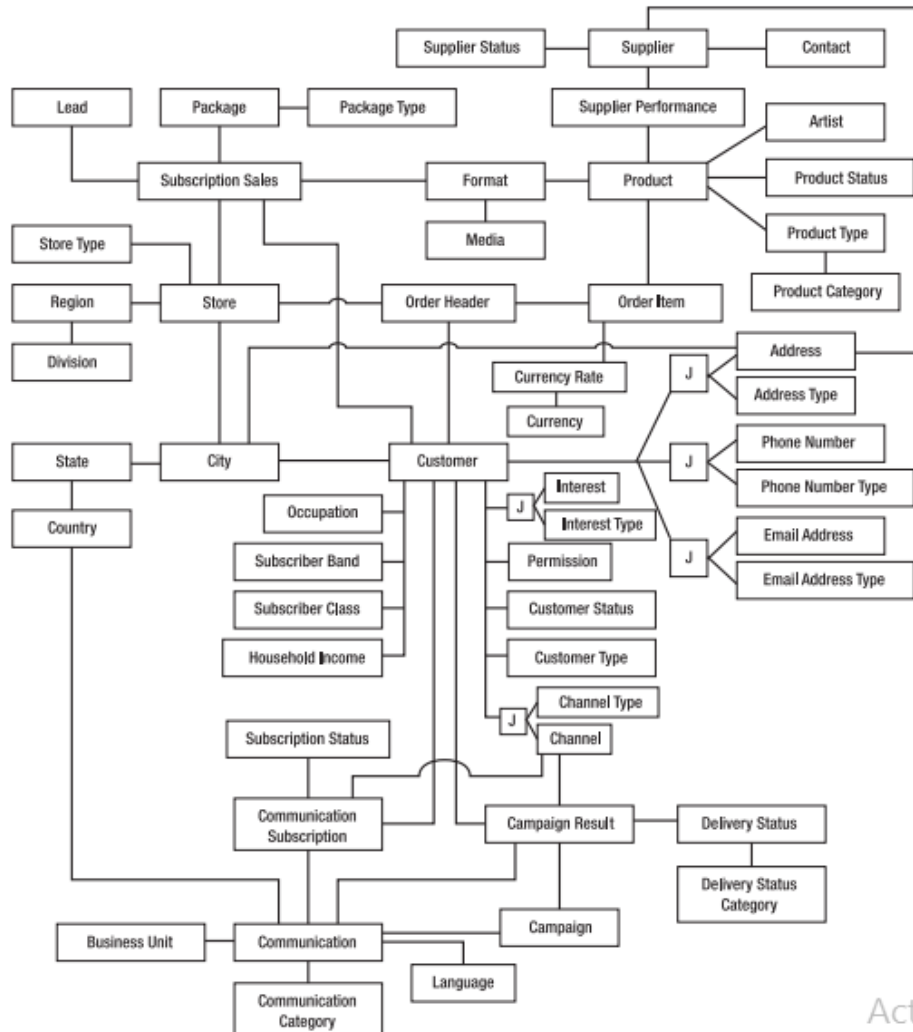
Thiết kế NDS

<u>Store</u>	<u>Product</u>	<u>Supplier</u>	<u>Customer</u>	<u>Order</u>	<u>Communication</u>
Store	Product	Supplier	Customer	Order Header	Interest
Store Type	Product Type	Supplier Status	Customer Status	Order Line	Permission
Web Site	Product Category	Contact	Address	Currency Rate	Channel
Region	Artist		State	Currency	Campaign
Division	Media	<u>Reference</u>	Country	Lead	Language
<u>Date</u>	Product Status	Customer Type	Phone Number	Source System	Communication
Fiscal Calendar	Package	Address Type	Email Address		Business Unit
Store Hours	Package Type	Phone Number Type	Occupation		Communication Category
Holiday	Format	Email Address Type	Household Income		Subscription Status
		Interest Type	Subscriber Class		Delivery Status
		Channel Type	Subscriber Band		Delivery Status Category

Activate Windows
Go to PC settings to activate Windows.



Thiết kế NDS





➔ Câu hỏi

- Nêu tư tưởng thiết kế vừa rồi (bottom up hay topdown,...)

➔ Ví dụ

- Subscription Sales fact table
 - Bảng Fact có 6 dimension:
 - lead, package, format, store, customer, **date**
 - Bảng Transaction nối với:
 - lead, package, format, store, customer
 - **Chiều ngày nên ánh xạ vào NDS như thế nào?**

➔ Gợi ý

Brief introduction to functions used in script to populate Date Dimension

• ht
9!
fo

	Function	Detail (e.g. for 16-Aug-2013)
1	Select DATEPART(MM, Getdate()) as MonthNumber	Return Integer Number=8 of Month from Current Date
2	Select DATEPART(YY , Getdate()) as YearValue	Return Value of the Year=2013 from Current Date
3	Select DATEPART(QQ , Getdate()) as QuarterValue	Return Value of the Quarter=3 for Current Date
4	Select DATEPART(DW, Getdate()) as DayOfWeekValue	Return integer Value of day=6 (Friday) in Week for Current Date as per US standard
5	Select CONVERT (char(8),Getdate(),112)	Return Key=20130816 Value for current Date
6	Select CONVERT (char(10),Getdate(),103)	Return date =16/08/2013 in "dd-MM-yyyy" format, UK, Europe
7	Select CONVERT (char(10),Getdate(),101)	Return date=08/16/2013 in "MM-dd-yyyy" format, US
8	Select DATEPART(DD , Getdate()) as DayOfMonthValue	Return integer Day=16 Value for Current Date
9	select DATENAME(DW, Getdate()) AS DayName	Return Name=Friday of the Day for Current Date.
10	select DATEPART(WW, Getdate()) AS WeekOfYear	Returns Value of Week in Year=33

647
n-

➡ ĐỒ ÁN LT

Chọn 1 trong 2

- Thiết kế data mart cho nhu cầu phân đoạn chiến dịch trong quản lý liên hệ khách hàng (CRM)
- Người dùng nghiệp vụ cần phân tích “bán hàng qua hình thức đăng ký” (subscription sales). Cần biết được doanh số, chi phí, lợi nhuận được đánh giá hằng ngày trong khoảng 1 tháng

An abstract graphic featuring a central point from which several colorful rays (orange, green, blue, grey) emanate, creating a starburst effect. The background is a light grey grid. Faint binary code (0s and 1s) is scattered across the grid. In the upper right, there is a small line graph with two lines, one orange and one blue, showing an upward trend. The text "Thank You!" is prominently displayed in the lower right area.

Thank You!

TB: Hồ Thị Hoàng Vy

2013, Jan 27