

# Seminar

**Môn : Phát Triển HTTT Theo Hướng Hiện Đại**

**Chủ Đề : Topic #20 - Tìm Hiểu Về DW 2.0  
(Chương 16,17,18)**

Nhóm: 18

MSSV	Họ Tên
1041309	Nguyễn Công Bình
1041317	Trần Ngọc Hoài Duy
1041440	Lương Minh Trí
1042122	Nguyễn Thị Thùy Trang

# Nội dung tìm hiểu

- Chương 16 : Di Trú
- Chương 17 : Biện Hộ Về Chi Phí và DW 2.0
- Chương 18 : Chất Lượng Dữ Liệu DW 2.0

# Chương 16 : Migration

Trình Bày : Trần Ngọc Hoài Duy

# Mục Tiêu của Migration

- Giúp chuyển đổi các data warehouse vào mô hình DW 2.0 một cách nhanh chóng và ít xảy ra lỗi.

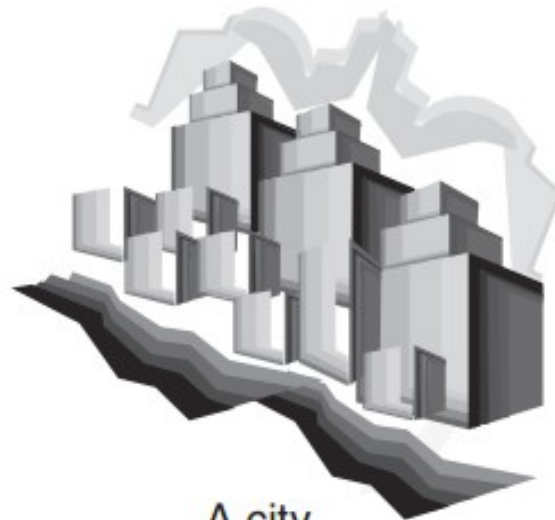
# Khái Niệm

So Sánh Giữa Xây Dựng :

Một ngôi nhà và một thành phố



A house



A city

When you are building DW 2.0, you are building a city, not a house.

# Xây Dựng một Ngôi Nhà

Một căn nhà được xây dựng trong một khoảng thời gian tương đối ngắn. Việc xây dựng một ngôi nhà luôn có thời điểm bắt đầu và kết thúc .

# Xây Dựng một Thành Phố

- Một thành phố có rất nhiều phần xây dựng khác nhau và được xây dựng trong một thời gian dài
  - Phần nào được xây dựng xong : sử dụng ngay
  - Có đặc điểm giống nhau giữa các thành phố nhưng cũng có những điểm riêng biệt

# Khái Niệm

- Việc thực thi DW 2.0 theo hướng khác nhau nhưng tất cả các các data warehouse DW2.0 vẫn chia sẻ cùng 1 kiến trúc
  - Nếu các công ty hầu như không bao giờ đặt ra xây dựng một môi trường DW 2.0 thì họ sẽ phát triển .
- => Công ty sẽ áp dụng Migration để di trú sang kiến trúc DW2.0



# Di trú trong một “thế giới Hoàn Hảo”

- “Thế giới hoàn hảo” bao gồm các thành phần :
  - Interactive Sector
  - Integrated Sector
  - Near Line Sector
  - Archival Sector

# Di trú trong một “thế giới Hoàn Hảo”

- Trong 1 “thế giới hoàn hảo” , thứ tự của việc xây dựng data warehouse DW 2.0 giống luồng của dữ liệu đi vào và thông qua môi trường DW 2.0.
- Data warehouse DW 2.0 được xây dựng trong trường hợp không tồn tại data warehouse nào trong cấu trúc của cơ sở hạ tầng.
- Khi mỗi cấp độ xử lý được xây dựng , thì sẽ làm nền tảng xây dựng cho cấp độ tiếp theo.

# “Thế giới hoàn hảo” hầu như không xảy ra

- “Thế giới hoàn hảo” chỉ là trình tự trên lý thuyết
- Data warehouse DW 2.0 hầu như không bao giờ được xây dựng trong các bước tuần tự từ trên xuống.
  - Lý do : đã tồn tại một môi trường data warehouse trước đó
- Những thành phần cơ bản nhất trong cơ sở hạ tầng :
  - Môi trường ứng dụng (Nguồn dữ liệu)
  - Tiến trình ETL (Khu vực xử lý)
  - CSDL hoặc data warehouse (Kho dữ liệu)

# Từng bước thực hiện thêm các thành phần

- Ưu thế của kiến trúc DW 2.0 là các thành phần được thêm vào tăng dần và độc lập. Giúp các công ty có thể di trú và phát triển môi trường DW 2.0 một cách có trật tự
- Nếu có thành phần không di trú đến kiến trúc DW 2.0 thì bị triệt tiêu và loại bỏ hệ thống hiện có.

# Từng bước thực hiện thêm các thành phần

- Mức lưu trữ Near – Line : là mức lưu trữ các truy cập và xử lý dữ liệu bởi người sử dụng
- Khi thêm near-line , nó giúp gia tăng tốc độ di trú đến các kiến trúc DW 2.0 , nó tạo ra môi trường data warehouse đầu tiên , tạo điều kiện dễ dàng để tạo được một kiến trúc để làm.
- Việc Attach 1 mức near-line mới , để có 1 data warehouse đầu tiên , thì không khó và cũng không cần phải chuẩn bị nhiều.

# Thêm Archival Sector

- Việc xây dựng Archival Sector cũng không cần được chuẩn bị trước.
- Archival Sector sẽ được nói rõ ở chương 19.

# Tạo Siêu Dữ Liệu Doanh Nghiệp

- Metadata - siêu dữ liệu (SDL) - là thông tin mô tả nội dung của cơ sở dữ liệu (CSDL). Metadata báo cho người dùng biết thời gian cập nhật sau cùng của dữ liệu, định dạng và mục đích sử dụng của nó. Những thông tin này có thể hướng dẫn người dùng duyệt qua CSDL và giúp họ hiểu được ý nghĩa và ngữ cảnh của dữ liệu tài chính, các bản ghi về khách hàng và các giao dịch kinh doanh.

Nguồn : <http://www.pcworld.com.vn/articles/cong-nghe/cong-nghe/2005/10/1184418/metadata-sieu-du-lieu/>

# Tạo Siêu Dữ Liệu Doanh Nghiệp

- Theo quy luật các metadata có được sử dụng hay không , nó vẫn được tạo ra .
- Các nhà cung cấp các công nghệ thường cung cấp một cơ sở để lưu trữ và quản lý metadata một cách dễ dàng như metadata ETL , metadata kinh doanh thông minh , và metadata DBMS
- Vì vậy , các nền tảng của metadata thường đã có.



# Tạo Siêu Dữ Liệu Doanh Nghiệp

- . Xây dựng metadata của doanh nghiệp thường gồm 3 bước sau :
    - Xây dựng kho lưu trữ metadata của doanh nghiệp.
    - Di dời các metadata đã có vào trong metadata của doanh nghiệp.
    - Biến đổi các metadata đã có cho phù hợp với định dạng của metadata doanh nghiệp
- => Bước cuối cùng là khó khăn nhất . bởi vì khó để biến đổi các metadata phù hợp với cấu trúc các công ty , doanh nghiệp .

# Xây dựng cơ sở hạ tầng Metadata

- Lúc trước , để xây dựng một kho lưu trữ metadata cấp doanh nghiệp , phải chỉnh sửa, cắt xén hoặc loại bỏ môi trường hiện tại.
- Ngược lại , cơ sở hạ tầng metadata DW 2.0 xây dựng được trên bất cứ cơ sở hạ tầng metadata đã có .

# “Nuốt” mã nguồn hệ thống

- Nếu có bất kỳ một hệ thống trong môi trường ứng dụng hiện hành ngừng hoạt động, thì đó là các ứng dụng cần phải được tiêu thụ bởi Interactive Sector của môi trường DW 2.0
  - Trong nhiều trường hợp , Interactive Sector sẽ “nuốt” những ứng dụng mã nguồn cũ .
  - Trong các trường hợp khác , thì những ứng dụng mã nguồn này sẽ được giữ lại và tiếp tục đóng góp dữ liệu cho Interactive Sector.

# “Nuốt” mã nguồn hệ thống

- Trong trường hợp ứng dụng mã nguồn bị “nuốt” bởi Interactive Sector , thì đó là 1 điều tốt cho những hệ thống đã lỗi thời.
- Những ứng dụng đã bị nuốt , được thiết kế để đáp ứng yêu cầu kinh doanh trong một thời gian dài.
- Nếu Interactive Sector không có tác động đến , thì những ứng dụng cũ phải được chỉnh sửa lại trong bất cứ trường hợp nào

# ETL như một thiết bị hấp thụ va chạm

- ETL là các công cụ trích xuất, chuyển đổi và nạp dữ liệu
- Khi thêm vào các ứng dụng mã nguồn , sẽ có những thay đổi lớn
- ETL hoạt động như một thiết bị hấp thụ va chạm cho việc phát triển toàn bộ data warehouse và tiến trình di trú.
- Thông qua việc chuyển đổi ETL trong Interactive Sector để xây dựng data warehouse một cách hiệu quả thì các thay đổi này được giảm thiểu hoặc có thể làm cho xấu hơn.

# Di trú trong môi trường không có cấu trúc

- Lĩnh vực dữ liệu không có cấu trúc là một trong những tính năng mới và quan trọng nhất của môi trường data warehouse DW 2.0
- Trong nhiều môi trường DW 2.0 , dữ liệu không có cấu trúc là thành phần được thêm vào để phát triển cho nhiều loại phân tích và xử lý hỗ trợ quyết định mới.
- Di trú sang môi trường DW 2.0 cho dữ liệu không có cấu trúc khá khác so với dữ liệu có cấu trúc

# Di trú trong môi trường không có cấu trúc

- Môi trường có cấu trúc hầu như luôn luôn tồn tại trong hình thức của một data warehouse ở thế hệ đầu tiên , trong khi đó , các thành phần không có cấu trúc thì không.
- Sau khi dữ liệu không có cấu trúc được xử lý để đưa vào data warehouse DW 2.0 , liên kết được hình thành giữa dữ liệu có cấu trúc và dữ liệu không có cấu trúc.
- Theo thời gian , dữ liệu không có cấu trúc không còn được sử dụng , nó được di trú vào lĩnh vực dữ liệu không có cấu trúc của Archival Sector DW 2.0 được giới thiệu ở chương 19.

