

Trường Đại học CNTT – Khoa Hệ thống thông tin

BÀI 3: KIẾN TRÚC KHO DỮ LIỆU

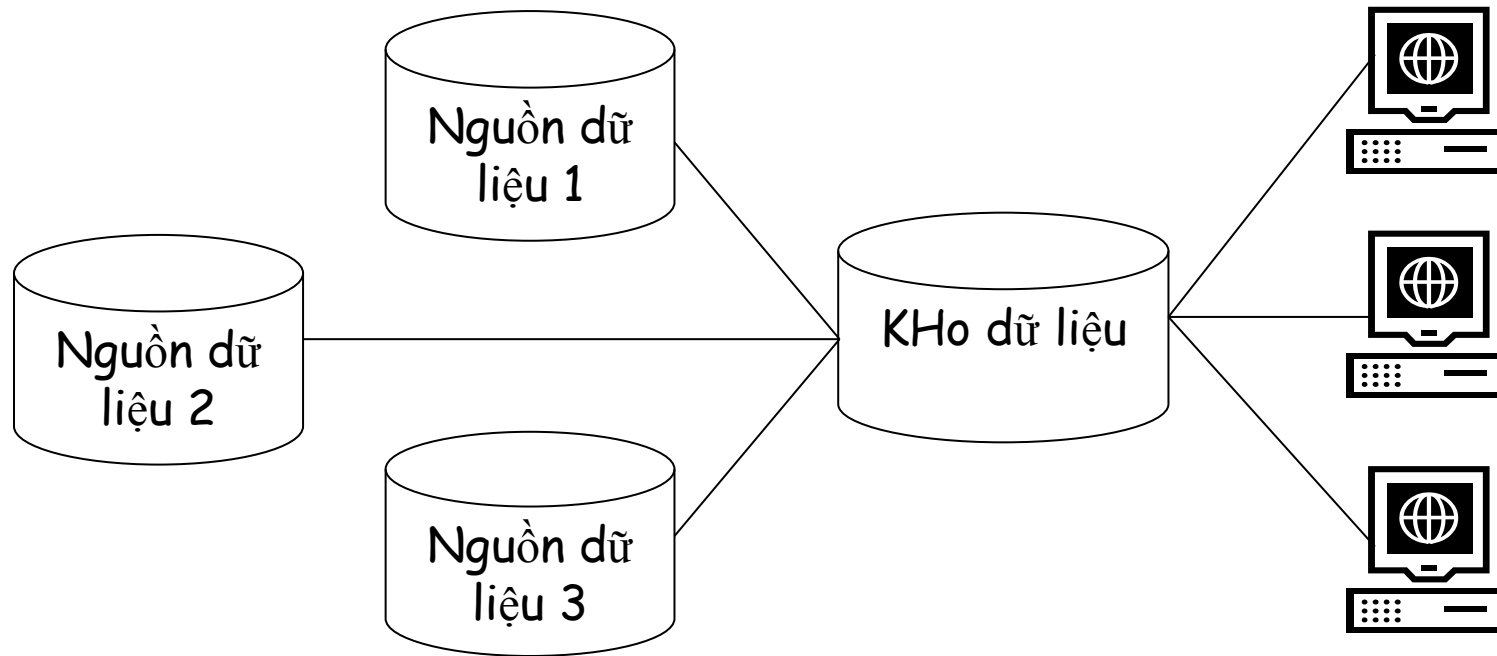
Giảng viên: ThS. Nguyễn Thị Kim Phụng
ThS. Hồ Trung Thành

Kiến trúc kho dữ liệu

- ▶ Kiến trúc 1 tầng
- ▶ Kiến trúc 2 tầng
- ▶ Kiến trúc 3 tầng
 - Nguồn dữ liệu
 - Kho dữ liệu
 - Các ứng dụng kho dữ liệu
- ▶ Hệ hỗ trợ ra quyết định

Kiến trúc 1 tầng

Môi trường tác vụ



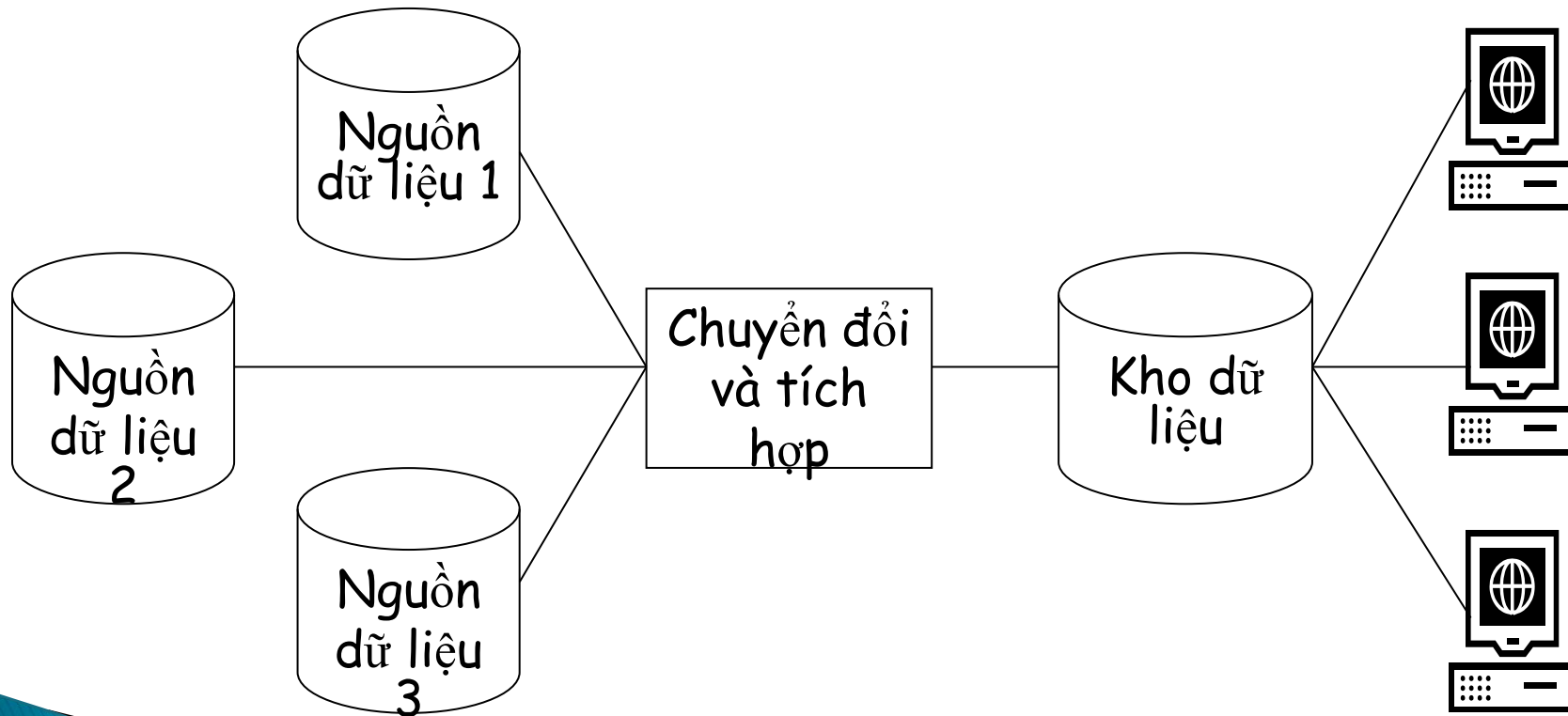
Kiến trúc 1 tầng (2)

- ▶ Người dùng cuối truy xuất dữ liệu từ các hệ thống xử lý nghiệp vụ thông qua kho dữ liệu
- ▶ Mức kho dữ liệu có nhiệm vụ tổng hợp dữ liệu từ các hệ thống xử lý nghiệp vụ mà không qua bước ETL

Kiến trúc 2 tầng

Môi trường tác vụ

Môi trường hỗ trợ
quyết định (Người
dùng cuối)



Kiến trúc 2 tầng

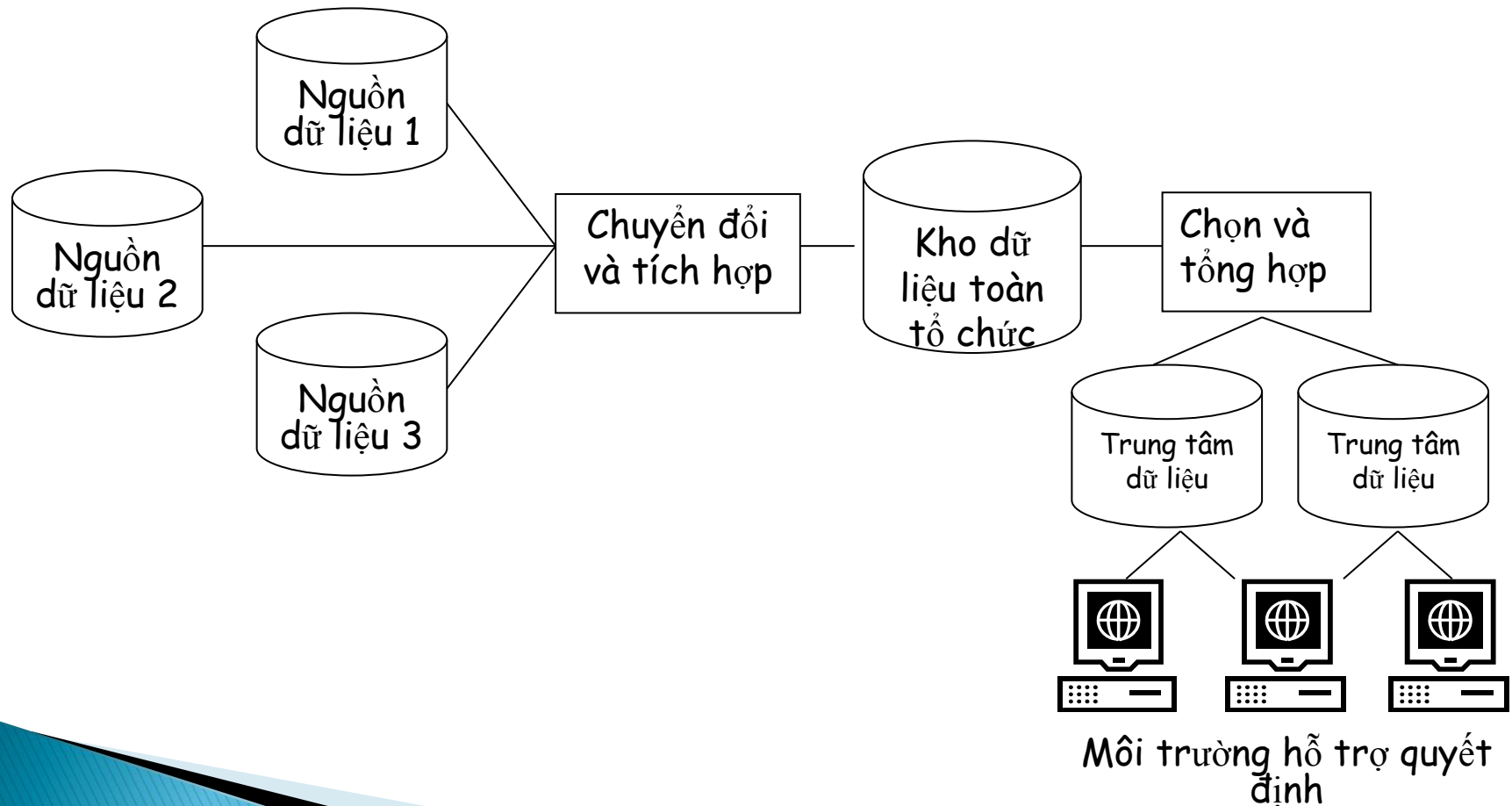
- ▶ Bước 1: Lấy dữ liệu từ các hệ thống nguồn khác nhau.
- ▶ Bước 2: Chuyển dạng và tích hợp các dữ liệu từ các hệ thống nguồn khác nhau trước khi đưa vào kho dữ liệu.
- ▶ Bước 3: Kho dữ liệu là một cơ sở dữ liệu chỉ đọc. Tổ chức kho dữ liệu để hỗ trợ quyết định. Kho dữ liệu chứa cả dữ liệu tổng hợp và dữ liệu chi tiết.
- ▶ Bước 4: Người sử dụng truy xuất kho dữ liệu bằng các phương tiện khác nhau như công cụ phân tích và ngôn ngữ truy vấn.

Kiến trúc 3 tầng

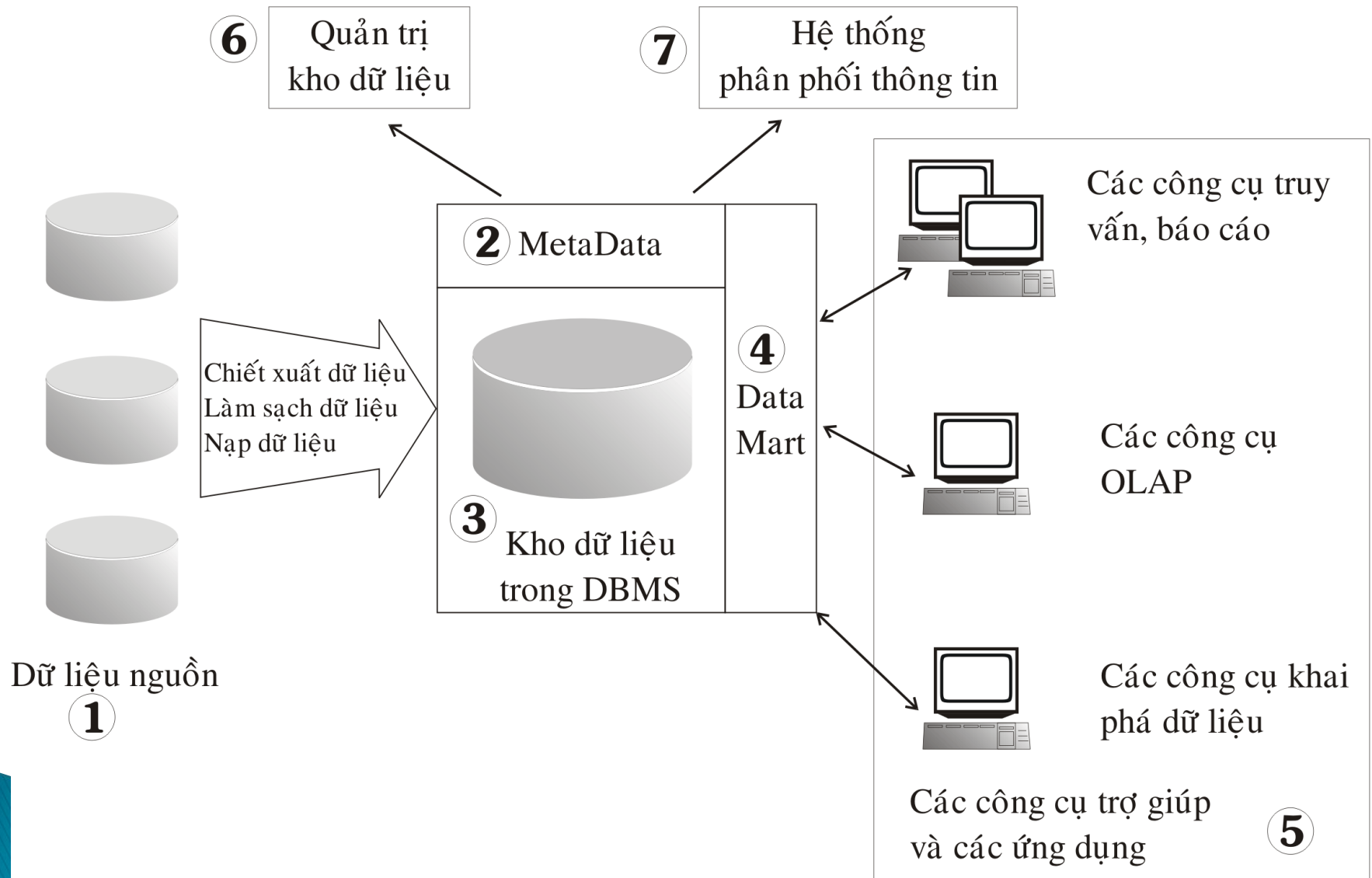
- ▶ Với kiến trúc ba tầng, dữ liệu được làm sạch, xử lý để đưa vào kho dữ liệu.
- ▶ Kho dữ liệu phân ra thành những kho dữ liệu theo chủ đề, đây là việc cần thiết cho những ứng dụng hỗ trợ ra quyết định theo từng nhóm người sử dụng.

Kiến trúc 3 tầng (2)

Môi trường tác vụ



Kiến trúc 3 tầng (3)



Kiến trúc 3 tầng (4)

- ▶ Với kiến trúc ba tầng, dữ liệu được ETL để đưa vào kho dữ liệu.
- ▶ Từ kho dữ liệu phân ra thành những kho dữ liệu theo chủ đề, đây là việc cần thiết cho những ứng dụng hỗ trợ ra quyết định theo từng nhóm người sử dụng

1. Nguồn dữ liệu (Data source)

- ▶ Dữ liệu nguồn (là các ứng dụng tác nghiệp hoặc các cơ sở dữ liệu tác nghiệp) và các công cụ chiết xuất, làm sạch và chuyển đổi dữ liệu.
- ▶ Nguồn dữ liệu có thể là một tập tin, hay một cơ sở dữ liệu của một HQTCSDL nào bất kỳ.

2. Siêu dữ liệu (Metadata)

- ▶ Siêu dữ liệu là tất cả những thông tin để miêu tả dữ liệu trong môi trường kho dữ liệu.
- ▶ Đối với nhóm kho dữ liệu, siêu dữ liệu gồm:
 - Thư mục về những gì có trong kho dữ liệu
 - Thư mục chỉ rõ nơi lưu trữ dữ liệu
 - Thư mục là chỉ mục dùng cho câu truy vấn thiên về việc tìm kiếm những thông tin đúng

2. Siêu dữ liệu (2)

- ❑ Những ánh xạ tương ứng giữa tác vụ và các thao tác trong kho dữ liệu.
- ❑ Dữ liệu được đưa vào kho dữ liệu phải ở dạng chuẩn và phải theo những quy ước của kho dữ liệu. Nghĩa là dữ liệu phải được làm sạch

2. Siêu dữ liệu (3)

- Những luật dùng để tổng hợp dữ liệu. Đối với người sử dụng kho dữ liệu, siêu dữ liệu gồm:
 - Những thuật ngữ dùng để miêu tả dữ liệu.
 - Những kỹ thuật tương ứng sử dụng để truy xuất dữ liệu.
- Nguồn dữ liệu, những luật dùng để suy dẫn nguồn dữ liệu và khi tạo ra nguồn dữ liệu

4. Kho dữ liệu theo chủ đề (Data mart)

- ▶ Data Mart là một dạng thu nhỏ của kho dữ liệu
- ▶ Nếu kho dữ liệu mô tả thông tin của một tổ chức thương mại thì:
 - Data Mart mô tả thông tin cho từng phòng ban của tổ chức đó (như phòng kinh doanh, phòng nhân sự,...)
 - Hoặc chứa thông tin cho mỗi chi nhánh của tổ chức.

4. Kho dữ liệu theo chủ đề (2)

- ▶ Data Mart đôi khi không thật sự cần thiết vì chúng có thể làm tốn thêm thời gian xử lý cho việc cập nhật dữ liệu mới cho một số lượng lớn Data Mart.
- ▶ Tuy nhiên, sử dụng Data Mart cho phép thực thi các truy vấn nhanh hơn vì dữ liệu của chúng ít hơn so với kho dữ liệu.

5. Các ứng dụng kho dữ liệu

- ▶ Một kho dữ liệu bình thường chưa đủ để có các tri thức. Để có tri thức, kho dữ liệu cần phải được khai thác và sử dụng đúng cách, có thể sử dụng theo 3 cách chính:
 - Theo cách khai thác truyền thống: các công cụ truy vấn, báo cáo
 - Xử lý phân tích trực tuyến
 - Khai phá dữ liệu

Hệ hỗ trợ ra quyết định (DSS – Decision Support System)

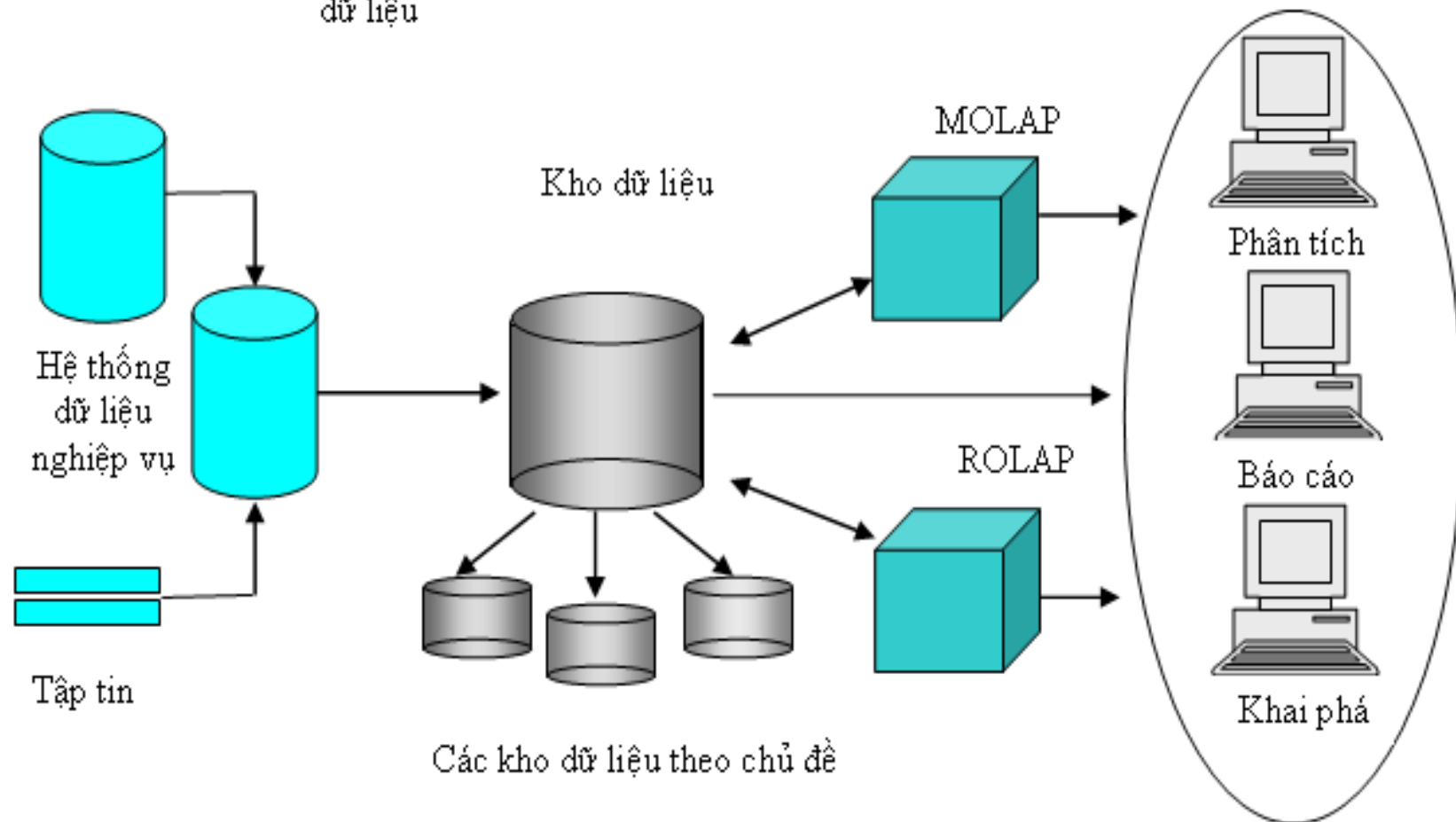
- ▶ Hệ hỗ trợ ra quyết định giúp những người lãnh đạo, quản lý và phân tích làm việc hiệu quả hơn
- ▶ Hỗ trợ cho việc ra quyết định trong quản lý, kinh doanh một cách hiệu quả và nhanh chóng

Hệ hỗ trợ ra quyết định (2)

- ▶ Hệ hỗ trợ ra quyết định mức tổng quát gồm 3 tầng:
 - Tầng 1: Kho dữ liệu.
 - Tầng 2: Hệ quản trị OLAP có thể là ROLAP, MOLAP , HOLAP
 - Tầng 3: Tầng người dùng hỗ trợ các công cụ truy vấn, báo cáo, phân tích, khai phá dữ liệu.

Hệ hỗ trợ ra quyết định (3)

Nguồn dữ liệu Vùng chuẩn bị dữ liệu Kho dữ liệu Hệ quản trị OLAP Người sử dụng



Danh mục từ viết tắt (1)

- ▶ Dimensions: các chiều
- ▶ Fact: sự kiện (dữ liệu sự kiện)
- ▶ Cube: khối dữ liệu
- ▶ Star-schema: lược đồ hình sao
- ▶ Snowflake-schema: lược đồ tuyết rơi
- ▶ Aggregation: tổng hợp
- ▶ Attribute hierarchies: các phân cấp thuộc tính
- ▶ Granularity: độ thô
- ▶ Metadata: siêu dữ liệu
- ▶ Report: báo cáo

Danh mục từ viết tắt (2)

- ▶ DWH: (Data Warehouse) Kho dữ liệu
- ▶ OLTP: (Online Transaction Processing):
Xử lý toàn tác trực tuyến
- ▶ OLAP: (Online Analytical Processing)
Xử lý phân tích trực tuyến
- ▶ Client/server OLAP: (client/server
Online Analytical Processing) Xử lý
phân tích trực tuyến khách chủ

Danh mục từ viết tắt (3)

- ▶ ROLAP: (Relational Online Analytical Processing) Xử lý phân tích trực tuyến quan hệ
- ▶ HOLAP: (Hybric Online Analytical Processing) Xử lý phân tích trực tuyến kết hợp
- ▶ DOLAP: (Database Online Analytical Processing) Xử lý phân tích trực tuyến CSDL
- ▶ MOLAP: (Multi dimensional Online Analytical Processing) Xử lý phân tích trực tuyến đa chiều