|  |
| --- |
| **ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**  **TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**  **KHOA: CÔNG NGHỆ THÔNG TIN**  **A blue and white logo  Description automatically generated**  **BÁO CÁO BÀI TẬP**  **So sánh kiến trúc Data Warehouses**  **MÔN: HỆ THỐNG THÔNG TIN**  **PHỤC VỤ TRÍ TUỆ KINH DOANH**    **NHÓM THỰC HIỆN – NHÓM 03:**  **MSSV: 20120049 – HỌ TÊN: Nguyễn Hải Đăng**  **MSSV: 20120113 – HỌ TÊN: Lê Nguyên Khang**  **MSSV: 20120269 – HỌ TÊN: Võ Văn Minh Đoàn**  **MSSV: 18120564 – HỌ TÊN: Lâm Hồng Thành**    **Giảng viên hướng dẫn: ThS. Hồ Thị Hoàng Vy**  **Lớp lý thuyết: 20\_1**  **Học kỳ - Niên khoá: HK1 - 2023-2024** |

MỤC LỤC

[I. Danh sách thành viên – Phân công bài tập 3](#_Toc147091383)

[II. Data Mart 3](#_Toc147091384)

[1. Giới thiệu 3](#_Toc147091385)

[2. Thành phần 4](#_Toc147091386)

[3. Ưu điểm 6](#_Toc147091387)

[4. Nhược điểm 6](#_Toc147091388)

[III. ODS 7](#_Toc147091389)

[1. Giới thiệu 7](#_Toc147091390)

[2. Thành phần 7](#_Toc147091391)

[3. Ưu điểm 8](#_Toc147091392)

[4. Nhược điểm 9](#_Toc147091393)

[IV. TÀI LIỆU THAM KHẢO 9](#_Toc147091394)

# Danh sách thành viên – Phân công bài tập

|  |  |  |
| --- | --- | --- |
| **Họ và tên** | **Mã số sinh viên** | **Phân công** |
| Nguyễn Hải Đăng | 20120049 | Data Mart – Nhược điểm |
| Lê Nguyên Khang | 20120113 | ODS – Nhược điểm |
| Võ Văn Minh Đoàn | 20120269 | Data Mart – Giới thiệu, Thành phần, ưu điểm |
| Lâm Hồng Thành | 18120564 | ODS – Giới thiệu, thành phần, ưu điểm |

# Data Mart

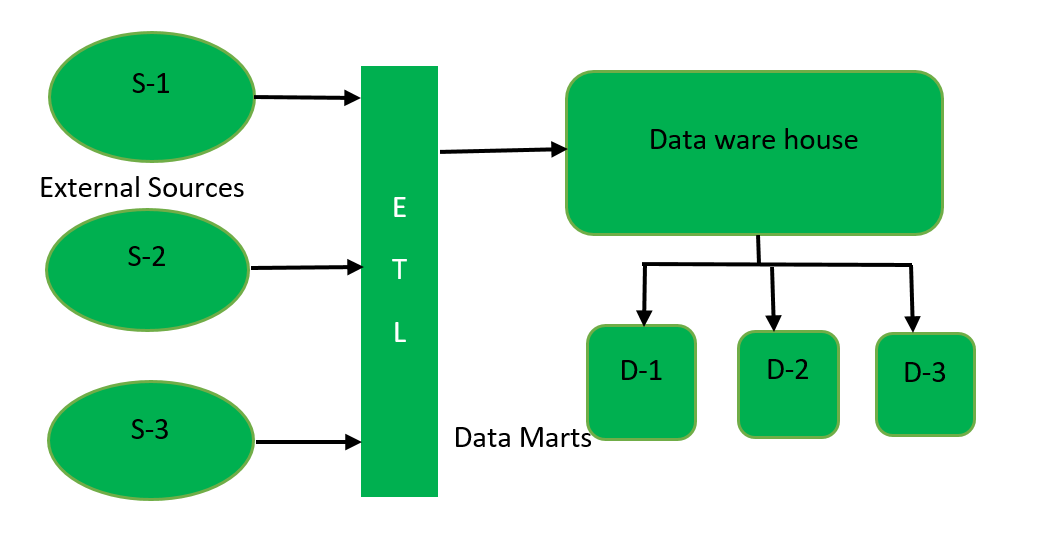
## Giới thiệu

* Data mart là 1 data warehouse nhỏ hơn được thiết kế xoay quanh một vấn đề, chức năng tổ chức, chủ đề hoặc lĩnh vực trọng tâm phù hợp khác.
* Data mart là 1 tập con của data warehouse và thường được định hướng cho một nhóm hoặc ngành kinh doanh cụ thể.
* Data mart về cơ bản là một phiên bản cô đọng và tập trung hơn của data warehouse phản ánh các quy định và thông số quy trình của từng đơn vị kinh doanh trong một tổ chức. Mỗi data mart được dành riêng cho một chức năng hoặc khu vực kinh doanh cụ thể. Tập hợp con dữ liệu này có thể trải rộng trên nhiều hoặc tất cả các lĩnh vực chức năng của doanh nghiệp. Thông thường, nhiều data mart sẽ được sử dụng để phục vụ nhu cầu của từng đơn vị kinh doanh riêng lẻ.
* Việc sử dụng cơ bản của data mart là các ứng dụng Business Intelligence (BI). BI được sử dụng để thu thập, lưu trữ, truy cập và phân tích bản ghi.
* So sánh data mart với data warehouse:
  + Data mart thường nắm giữ 1 lĩnh vực chủ đề (tài chính, bán hàng,...); data warehouse nắm giữ nhiều lĩnh vực chủ đề.
  + Data mart có thể chứa nhiều dữ liệu tóm tắt hơn; data warehouse chứa dữ liệu rất chi tiết.
  + Data mart tập trung vào việc tích hợp thông tin từ một lĩnh vực chủ đề nhất định hoặc một tập hợp các hệ thống nguồn; data warehouse hoạt động để tích hợp tất cả các nguồn dữ liệu.
  + Data mart được xây dựng tập trung vào dimensional model bằng cách sử dụng lược đồ sao; data warehouse không nhất thiết phải sử dụng mô hình thứ nguyên nhưng cung cấp dimensional model.

## Thành phần

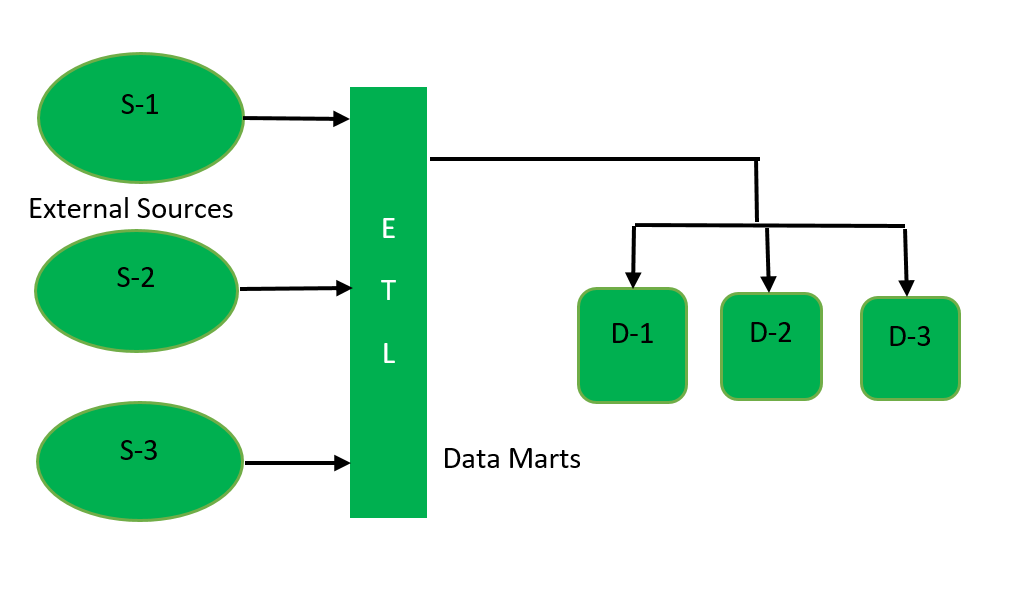
Có 3 cách thiết kế Data Mart:

* Dependent data mart: data mart được tạo ra bằng cách lấy dữ liệu từ data warehouse và phụ thuộc vào data warehouse.



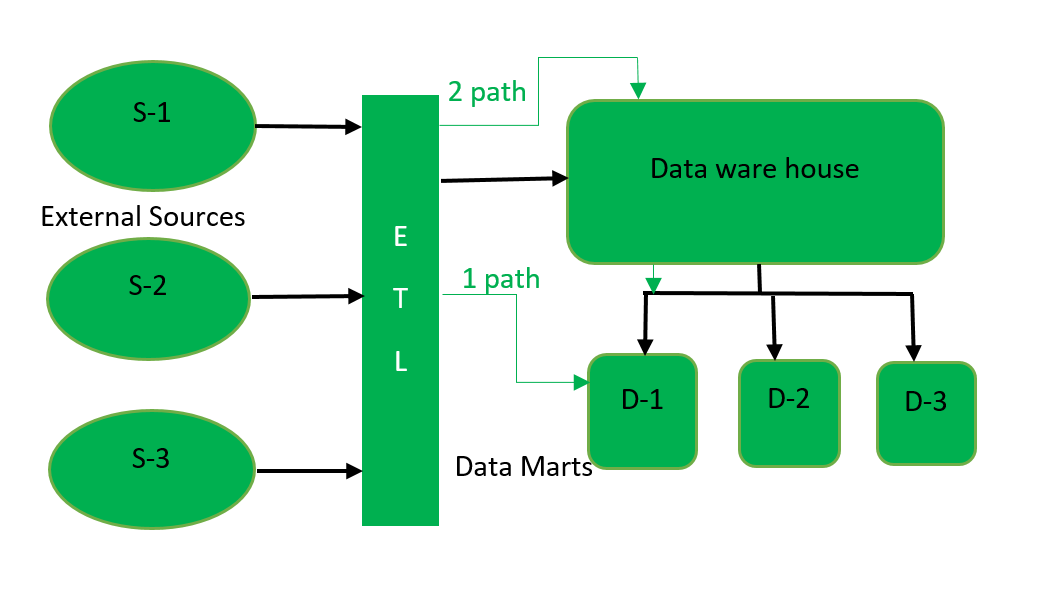
Dependent Data Mart được tạo bằng cách trích xuất dữ liệu từ kho lưu trữ trung tâm, Data warehouse. Kho dữ liệu đầu tiên được tạo bằng cách trích xuất dữ liệu (thông qua công cụ ETL) từ các nguồn bên ngoài và sau đó data mart được tạo từ data warehouse. Dependent data mart được tạo theo cách tiếp cận từ trên xuống của kiến trúc data warehouse. Mô hình data mart này được các tổ chức lớn sử dụng.

* Independent data mart: datamart được tạo ra từ nguồn bên ngoài mà không sử dụng data warehouse và data mart độc lập với data warehouse.



Independent Data Mart được tạo trực tiếp từ nguồn bên ngoài thay vì data warehouse. Data mart đầu tiên được tạo bằng cách trích xuất dữ liệu từ các nguồn bên ngoài và sau đó data warehouse được tạo từ dữ liệu có trong data mart. Siêu thị dữ liệu độc lập được thiết kế theo cách tiếp cận từ dưới lên của kiến trúc kho dữ liệu. Mô hình data mart này được các tổ chức nhỏ sử dụng và tương đối hiệu quả về mặt chi phí.

* Hybrid data mart: data mart được tạo ra từ nguồn vận hành hoặc data warehouse.



Loại Data Mart này được tạo bằng cách trích xuất dữ liệu từ nguồn vận hành hoặc từ data warehouse. 1Path phản ánh việc truy cập dữ liệu trực tiếp từ các nguồn bên ngoài và 2Path phản ánh mô hình dữ liệu phụ thuộc của data mart.

## Ưu điểm

* Việc triển khai data mart cần ít thời gian hơn so với việc triển khai datawarehouse vì data mart được thiết kế cho một bộ phận cụ thể của một tổ chức.
* Data mart là giải pháp thay thế tiết kiệm chi phí hơn so với data warehouse.
* Các tổ chức được cung cấp các lựa chọn để lựa chọn mô hình data mart tùy thuộc vào chi phí và hoạt động kinh doanh của họ.
* Dữ liệu có thể dễ dàng truy cập từ data mart.
* Data mart rất dễ sử dụng vì nó được thiết kế đặc biệt cho nhu cầu của người dùng. Do đó, siêu thị dữ liệu có thể đẩy nhanh các quy trình kinh doanh.
* Nó chứa các truy vấn được truy cập thường xuyên, vì vậy cho phép phân tích xu hướng kinh doanh.
* Bảo mật được tách biệt: Vì data mart chỉ chứa dữ liệu cụ thể cho bộ phận đó, bạn được đảm bảo rằng không thể truy cập dữ liệu ngoài ý muốn (dữ liệu tài chính, dữ liệu doanh thu).
* Hiệu suất được tách biệt: Vì mỗi data mart chỉ được sử dụng cho một bộ phận cụ thể, tải hiệu suất được quản lý và truyền đạt tốt trong bộ phận, do đó không ảnh hưởng đến các khối lượng công việc phân tích khác.

## Nhược điểm

* Vì nó chỉ lưu trữ dữ liệu liên quan đến chức năng, bộ phận cụ thể nên không lưu trữ khối lượng dữ liệu khổng lồ liên quan đến từng bộ phận của một tổ chức như Data Warehouses. Khả năng phân tích và báo cáo vì thế cũng bị giới hạn.
* Việc tạo quá nhiều Data Mart đôi khi trở nên cồng kềnh.
* Tính linh hoạt hạn chế: Data Mart bị giới hạn bởi dữ liệu và cấu trúc có sẵn trong Data Warehouses.
* Hiệu suất: Các Data Mart có thể cập nhật chậm hơn vì chúng dựa vào Data Warehouses để cập nhật.
* Đôi khi bạn có thể có thể nhận được rất nhiều dữ liệu vô Data Mart nhưng chúng không liên quan gì tới nhau. Điều này sẽ gây cản trở rất lớn tới việc bảo trì.

# ODS

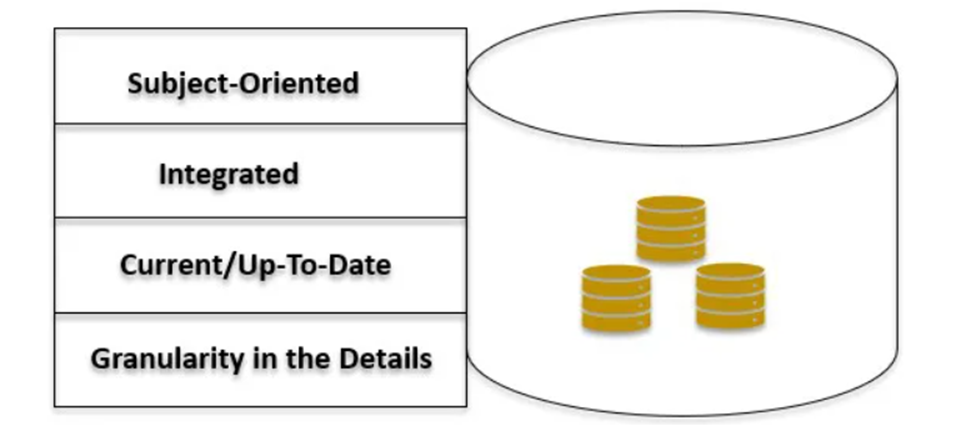
## Giới thiệu

* ODS (Operation Data Store): là một thành phần của data architecture.
* Mục tiêu chính của ODS là tổng hợp dữ liệu từ nhiều hệ thống nguồn càng gần thời gian thực càng tốt để hỗ trợ các xử lý kinh doanh cụ thể hoặc báo cáo vận hành.
* Sự khác biệt của ODS so với DW:
  + Tải dữ liệu về một nguồn duy nhất, dữ liệu trong ODS chưa được cleaning, transforming như trong DW.
  + Chỉ tải dữ liệu hiện tại chứ không lưu lại dữ liệu quá khứ.
  + Chỉ tải những dữ liệu cụ thể có ích cho quy trình kinh doanh.

## Thành phần

Khi tạo operational data store, nhiều nguồn dữ liệu có thể được tích hợp. Tuy nhiên, mỗi hệ thống nguồn dữ liệu cần phải đảm bảo các nguyên tắc sau:

* Subject-Oriented: Thiết kế của Kho dữ liệu vận hành phải được xây dựng dựa trên các yêu cầu chức năng của doanh nghiệp, đặc biệt là liên quan đến một lĩnh vực cụ thể đang được quan tâm.
* Integrated: Tất cả dữ liệu từ nhiều nguồn khác nhau phải trải qua quy trình ETL, bao gồm việc chuyển đổi dữ liệu thành một định dạng duy nhất và tải tập dữ liệu vào ODS.
* Up-to-Date: Dữ liệu ODS phải cập nhật và do đó được cập nhật để lưu trữ tất cả các chức năng gần đây của ứng dụng được kết nối với kho dữ liệu, cũng như để mô tả trạng thái hiện có của dữ liệu.
* Detailed: Khi các quy tắc được triển khai, điều quan trọng là phải duy trì mức độ chi tiết toàn diện của doanh nghiệp để thực hiện đúng các chức năng tương ứng, chủ yếu hỗ trợ các yêu cầu hoặc chức năng vận hành kinh doanh.



## Ưu điểm

* ODS cung cấp nhiều khả năng truy cập hơn vào dữ liệu vận hành và xử lý truy vấn không phức tạp trên một lượng nhỏ dữ liệu một cách nhanh gọn.
* ODS cung cấp thông tin cho việc ra quyết định dựa trên dữ liệu hiện tại, gần thời gian thực.
* ODS tích hợp dữ liệu từ các hệ thống mới và hiện có, tạo ra kho lưu trữ dữ liệu trung tâm.
* Các tổ chức có cái nhìn đầy đủ hơn về các chức năng kinh doanh khác nhau và cái nhìn trạng thái hiện tại giúp xác định, chẩn đoán các vấn đề dễ dàng hơn.
* Doanh nghiệp có thể xây dựng các quy tắc trên ODS để những thay đổi về dữ liệu trong một hệ thống cụ thể sẽ kích hoạt hành động trên một hệ thống khác.
* ODS có khả năng phục hồi tốt hơn trước các cuộc tấn công mạng.
* ODS có thể hoạt động như một giai đoạn trung gian trước khi dữ liệu được đưa vào kho dữ liệu. Nó có thể giúp dữ liệu trở nên nhất quán và do đó cải thiện chất lượng dữ liệu trong kho dữ liệu.
* ODS thường có thể mở rộng quy mô dễ dàng vì nó không cần hỗ trợ các quy trình ETL phức tạp.
* ODS có thể dễ dàng điều chỉnh các thay đổi về yêu cầu hoặc nguồn dữ liệu hơn do kiến trúc đơn giản hóa của nó.

## Nhược điểm

* Khó mở rộng: ODS không linh hoạt trong việc thích nghi với sự thay đổi trong nhu cầu sử dụng dữ liệu và phân tích.
* Độ trễ cao: ODS chỉ thích hợp cho những tập dữ liệu nhỏ. Khi query từ các tập dữ liệu lớn, ODS có năng suất rất thấp, có thể gây ảnh hưởng tiêu cực đến doanh nghiệp.
* Khả năng đồng thời kém: Hiệu suất có ODS sẽ giảm đáng kể nếu có nhiều user truy xuất đồng thời.

# TÀI LIỆU THAM KHẢO

<https://datafinz.com/benefits-of-operational-data-store/>

<https://www.starthubpost.com/operational-data-store-advantages-and-disadvantages/>

<https://hevodata.com/learn/operational-data-store-a-comprehensive-guide/>

<https://www.geeksforgeeks.org/data-marts-storage-component-of-hdfs/>

<https://www.guru99.com/data-mart-tutorial.html>

<https://en.wikipedia.org/wiki/Data_mart>

<https://viblo.asia/p/data-mart-la-gi-tim-hieu-tong-quan-ve-data-mart-GyZJZn0kJjm>

<https://streamsets.com/blog/operational-data-stores-ods-data-warehouses/>

<https://www.fivetran.com/learn/data-mart>

<http://www.rapid-business-intelligence-success.com/data-mart.html>

Data architecture.pdf (tài liệu của cô Vy)