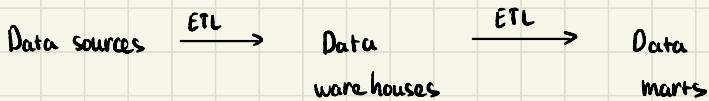


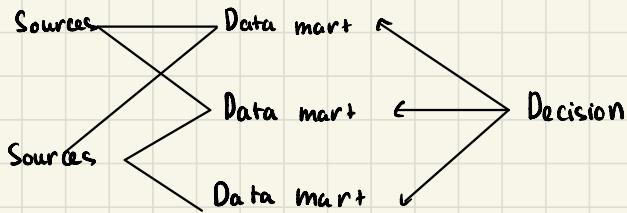
Compare a data warehouse to a data mart



Dependent data marts

1. No data warehouse no data marts because they can be supplied with data

Independent data marts



1. Independent data mart doesn't need a data warehouse
2. Each independent data mart draws data directly from one or more source

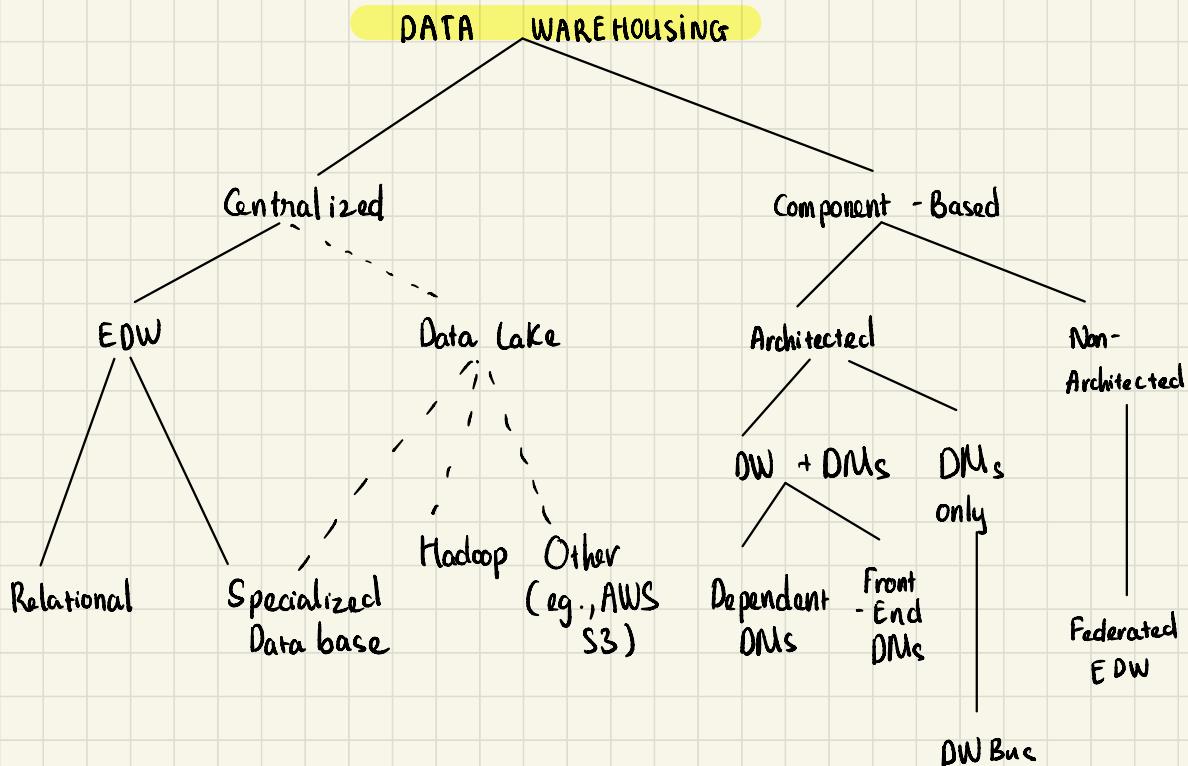
application

Dependent	Independent
Sourced from data warehouse	Sources directly from applications and system
(Mostly) uniform data across marts marts thông nhât	Little or no uniformity across marts ⇒ thông nhât
Architecturally straight forward	"Spaghetti" architecture

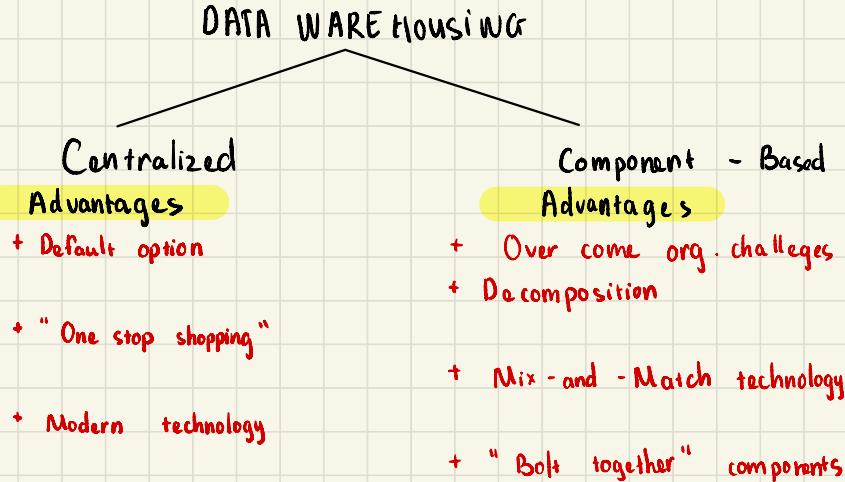
Data warehouse	Independent data mart
Many sources	One or more source
ETL from sources	ETL from sources
Probably largest data volumes	Possibly large data volumes
Dimensionally organized data	Dimensionally organized data

Your Data Warehousing Architectural Options

Many architectural options



1. First decision



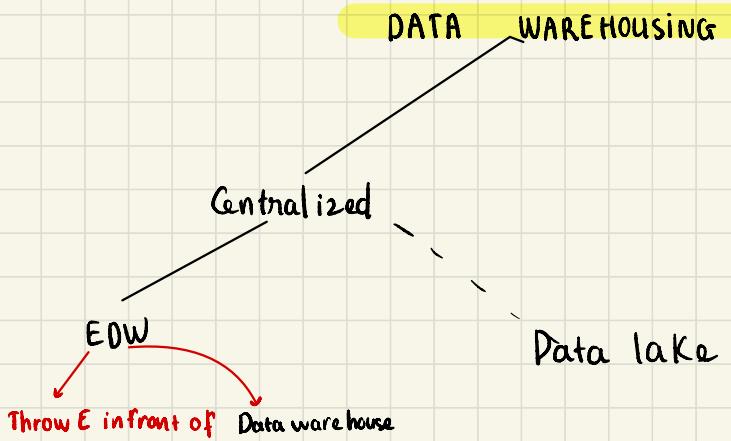
Disadvantages

- + High cross-org cooperation
(Yêu cầu hợp tác tổ chức cao)
- + High data governance
(Yêu cầu cao về quy tắc)
- + Ripple effects
(Hiệu ứng lướt sóng ,
khi có 1 thay đổi nhỏ dùn
ra thì sẽ ảnh hưởng tổng
thể)

Disadvantages

- + Often inconsistent data
- + Difficult to cross - integrate
(tích hợp chéo)

2. Emphasis on "enterprise"



=> We can think of an enterprise data warehouse (Kho dữ liệu doanh nghiệp)

as the default approach when we're building centralized environment

EDW

Relational

Specialized
Database

Specialized large scale databases and other things that are typically known as Warehousing applications (thiết bị lưu trữ dữ liệu)

Data Lake

Specialized
Data base

Hadoop

Other (e.g., AWS
S3)

Công nghệ lưu trữ

DATA WAREHOUSING

Component - Based (Hubing thành phần)

Architected

(có cấu trúc)

Non-
Architected

(không có cấu trúc)

DW + DMs

DMs Only

(Data warehouse + Data Marts) (Data Marts only)

Dependent
DMs

Front-end
DMs

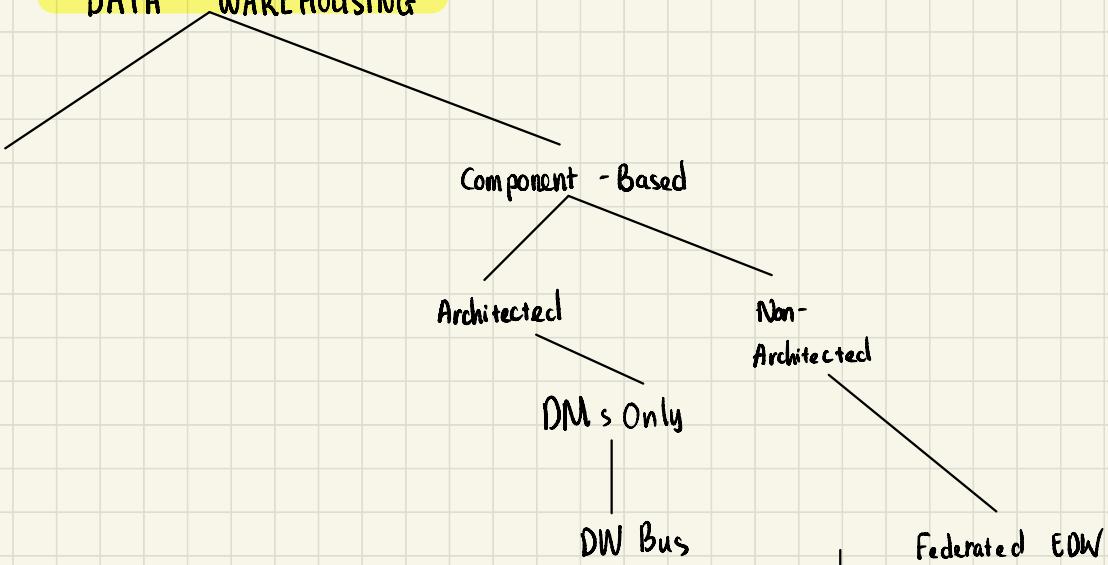
Data Sources $\xrightarrow{\text{ETL}}$ Data warehouse $\xrightarrow{\text{ETL}}$ Data Marts

bên thứ 3
Variation of Dependent DMs call

Data Marts \rightarrow Data Warehouse

Corporate information factory or CIF
(nhà máy thông tin)

DATA WAREHOUSING



Data marts follow a principle

that's known as **conformed dimensions**

(Thứ nguyên phù hợp)

⇒ Phương pháp tổ chức kho dữ liệu mà trong đó có các quy trình nghiệp vụ và các thuộc do xác định 1 cách rõ ràng và đồng nhất

Federated EDW
Kho dữ liệu liên hợp

o thường nhất về quy tắc kinh doanh, các mô hình và cấu trúc dữ liệu, cũng như mọi thứ cần thiết để xây dựng 1 kho dữ liệu tập trung hoặc datawarehouse bus

built a collection of independent data marts

What is a cube?

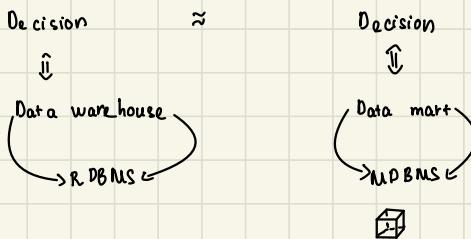
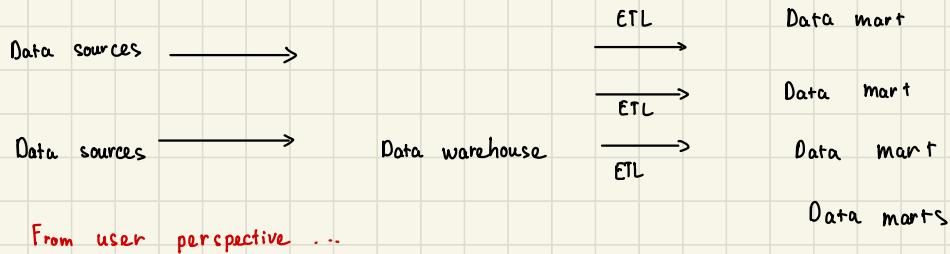
- * Cube = Multi dimensional database (MDDBMS)
- * Not a relational data base (RDBMS)
- * Specialized "dimensionally-aware" database

Today : best for smaller-scale DWs, DMs

CUBE : advantages and disadvantages

1. Fast query response time
2. "Modest" data volumes (khảm tốn)
3. Less flexible data structures than RDBMS

Data warehouses and marts together



Including Operational Data Stores in Your Data Warehousing environment

What is the difference between a DW and an ODS?

What is ODS?

- * Integrates data from multiple source
- * Emphasis on current operational data

Nhận mãnh

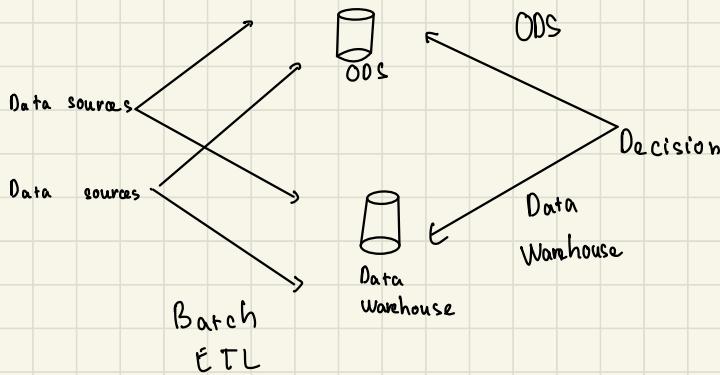
- * Often real-time source → ODS data feeds

→ Thực tiếp truyền data horn
thay vì xop hàng đợi datawarehouse refresh

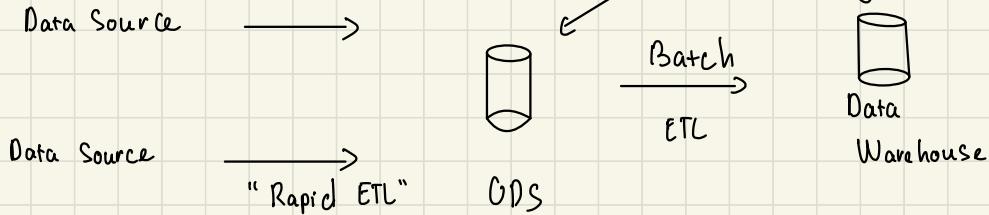
- * "Tell me what happening right now"

- * Popular late 1990s / early 2000s

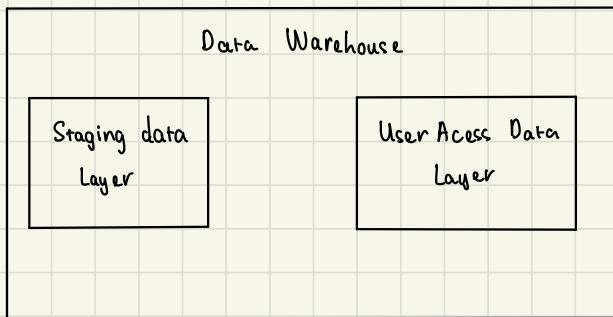
ODS and warehouses : Option 1



Option 2



Role of staging Layer



Staging Layer

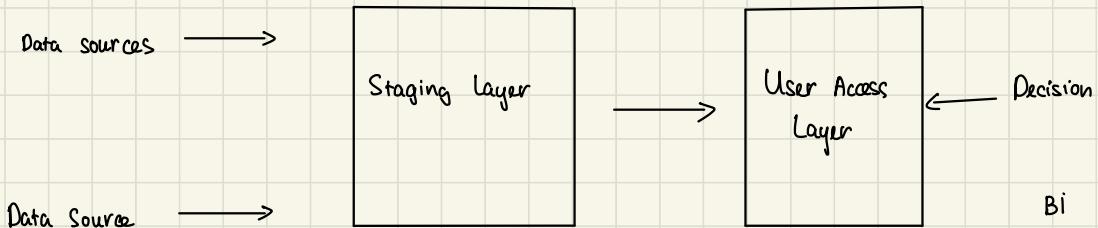
* "Landing Zone"

- "E" within ETL
- 2 variations

User access layer

- Where user go
- Dimensional data

Expanding our data warehousing architecture



In side staging area

