MOTIVATION

Operational processing: captures, stores and manipulates data to support daily operations.

Information processing: analysis of data to support decision making = BD2
└→ process of collecting and analysis different types of data

→ pekiştirmek

Data Warehouse: it can consolidate and integrate information from internal and external sources and arrange it in a meaningfull format for making business decision making.

DEFINITION

Data Warehouse: ilişkili verilerin sorgulandığı ve analizlerin yapıldığı bir depodur.
- veri tabanını yormamak için oluşturulmuştur.
- yalnızca sorgulamak ve analiz yapmak amacıyla kullanılır. → microsoft analysis services
- geçmişe ait büyük miktarlarda veri içerir.
⇒ veriyi kolay, hızlı ve doğru biçimde analiz etmek için gerekli işlemleri yapar

Data Warehousing: process of constructing and using DW

Data Warehouse

Subject-Oriented
- Organized around major subjects

- Focusing on the modelling and analysis of data for decision makers.

- Provide a simple and concise view
  özlü görüş

Integrated
- Constructed by integrating multiple, heterogeneous data sources.

- Data cleaning and data integration techniques are applied.

- When data is moved to the DW it is converted.

Time-variant
- Time horizon for the dw is longer than of OS

- Provide info from a historical perspective

- Every key structure in the DW contains
  → element of time, explicit or implicity
  açıkça   örtük

Non Updatable
- Operational update of data does not occur in the DW environment.

- Requires two operations in data accessing:
  → initial loading of data
    insert
  → acces of data
    select

DW ARCHITECTURES

Generic two-level architecture

- client-server application
- direct communication between client and data source server → data layer / database layer

Client applications
Data source
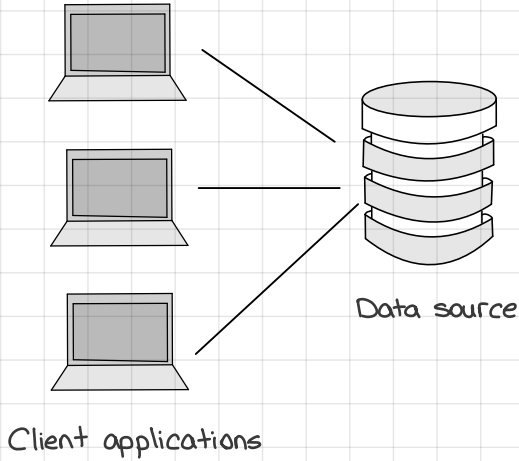
- directly connected to data sources layer without any intermediate applications

advantages
- easy to maintain
- modification of the stored data is easy

disatvantages
- performans will be degraded with increase user traffic
- cost-ineffective

Independent Data Mart

Data Mart: Focused on a single functional area of an organization and contains a subset of data stored in Data Warehouse.

Bir organizasyonun tek bir işlevsel alanda odaklanmış ve bir veri ambarına (DW) depolanan veri alt kümesi içerir.


Operational sources
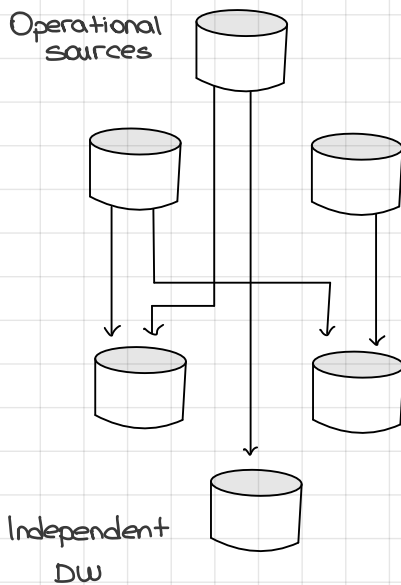
Independent DW

- independent DM is created without the use of Central DW
- ideal for smaller groups within an organization
- IDM has neither a relationship with the enterprise DW nor with any other DW. → No relationship with DW
- Data is input seperately

⇒ Independent Data Mart: a data mart filled with data extracted from the operational enviroment without benefits of DW.

Veri ambarının (DW) faydaları olmadan operasyonel ortamdan çıkarılan verilerle dolu bir veri pazarı.

-draws data from a central DW that has already been created
↳ önceden oluşturulmuş merkezi bir veri ambarından veri alır.

Operational sources

Enterprise DW

Dependent departmental Dm
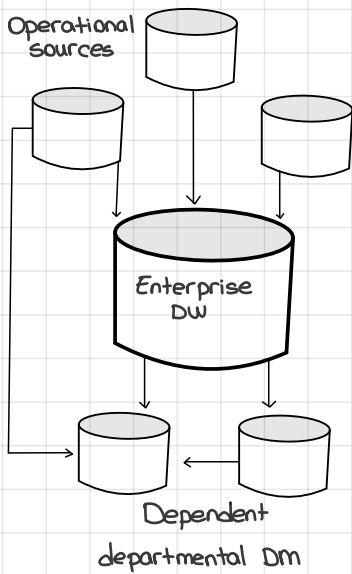
- DDM: A Dm filled exclusively from the enterprise DW and its reconciled data.
- Operational Data Store (ODS): an integrated, subject-oriented, updatable, current-valued, enterprise-wise, detailed database designed to serve operational users as they do decision support processing

Active Data Warehouse

- an interprise DW that accepts near-real-time feeds of transactional data from the systems of record, analyzes warehouse data, and in near-real-time-relays business rules to the DW and systems of record so that immediate actions can be taken in response to business events.

Data Warehouse	Data Mart
Scope • application implemented • centralized, possibly enterprise-wide • planned	**Scope** • specific DDS application • decentralized by user area • organic, possibly not planned
Data • historical, detailed and summarized • lightly denormalized	**Data** • some history, detailed and summarized • highly denormalized
Subjects • multiple subjects	**Subjects** • one central subject of concern to users
Sources • many internal and external sources	**Sources** • few internal and external sources
Other characteristic • flexible • data-oriented • long life • large • single complex structure	**Other characteristic** • restrictive • project-oriented • short life • starts small becomes large • multi, semi-complex structure, together complex

enterprise data model

Data mart — Derived data — Data Mart metadata

Reconsiled data — Enterprise DW and operational data store — EDW metadata

Operational data — Operational systems — operational metadata
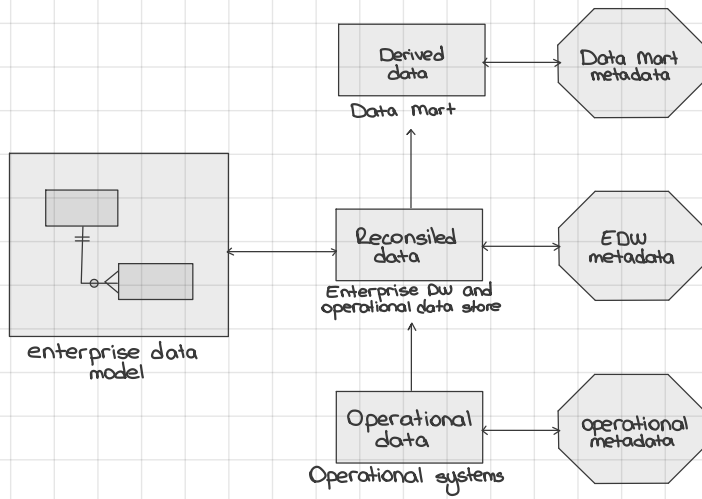
- Seperates its tier from each other based on the complexity of the users and how they used the data present in the DB.

katmanlarını, kullanıcıların karmaşıklığına ve veri tabanında bulunan verileri nasıl kullanıldıklarına göre birbirinden ayırır.

reconciled data: detailed, current data intended to be the single, authoritative source for all decision support.

derived data: data that have been selected, formatted and aggregated for end-user decision support application.

metadata: technical and business data that describe the properties or characteristics of other data.

DATA RECONCILIATION

- Typical Operational data is:
  - transient → not historical
  - not normalized (perhaps due to denormalization for performance)
  - restricted in scope → not comprehensive
  - sometimes poor quality → inconsistencias and errors

- After ETL, data should be:
  - detailed → not summerized yet
  - historical - periodic
  - normalized - 3rd normal form or higher
  - comprehensive → enterprise-wide perspective
  - quality controlled → accurate with full integrity