

# Architetture di Data Warehouse

---



# L'aspetto architettonale del DW

---

- L'aspetto architettonale è estremamente importante: il dw deve essere dotato della capacità di eseguire operazioni complesse e 'pesanti' dal punto di vista dell'impegno di capacità di elaborazione; è quindi richiesta una significativa capacità di progettazione e pianificazione dell'architettura di supporto del sistema.

## Kelly (1999) definisce come irrinunciabili per un sistema di data warehouse le caratteristiche di:

---

- q **Separazione:** l'elaborazione analitica e quella operativa devono essere mantenute il più possibile separate;
  - q **Scalabilità:** l'architettura hardware e software deve poter essere facilmente ridimensionata a fronte della crescita nel tempo dei volumi di dati da gestire ed elaborare e del numero di utenti da soddisfare;
  - q **Estendibilità:** deve essere possibile accogliere nuove applicazioni e tecnologie senza riprogettare integralmente il sistema;
  - q **Sicurezza:** il controllo sugli accessi è essenziale a causa della natura strategica dei dati memorizzati;
  - q **Amministrabilità:** la complessità dell'attività di amministrazione non deve risultare eccessiva;
-

# L'architettura

---

**Esistono diverse architetture tipiche dei sistemi di data warehouse, tra le quali ricordiamo:**

- A un livello**
- A due livelli**
- A due livelli con data-mart indipendenti**
- A tre livelli**

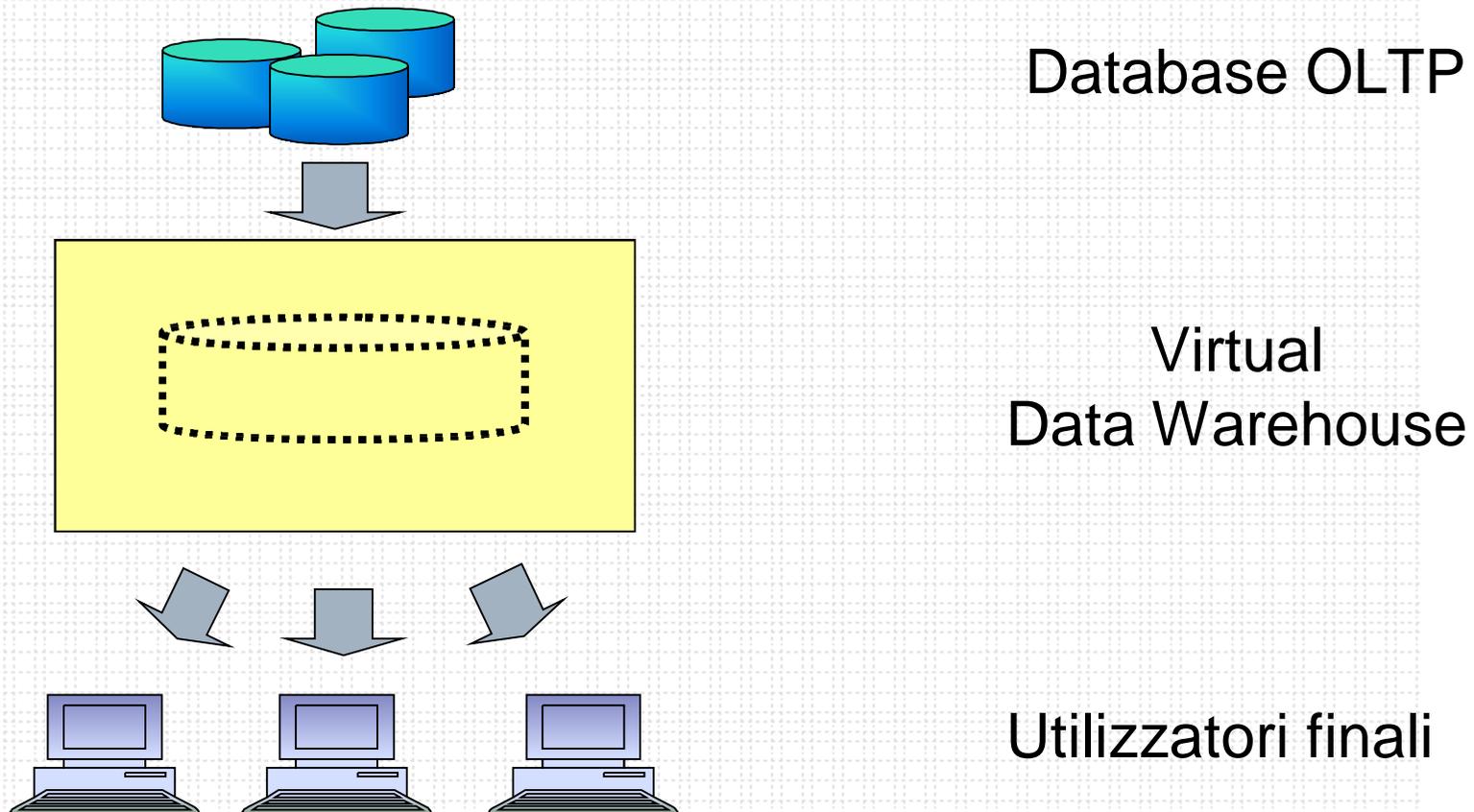
# A un livello

---

- Obiettivo di questa architettura è la minimizzazione dei dati memorizzati, ottenuta eliminando le ridondanze.
  - Questo DW è virtuale, nel senso che viene implementato come una vista multidimensionale dei dati operazionali generata da un apposito “middleware”, ossia da uno strato d’elaborazione intermedio (Devlin, 1997).
-

# Virtual Data Warehouse: a un livello

---



# Problemi

---

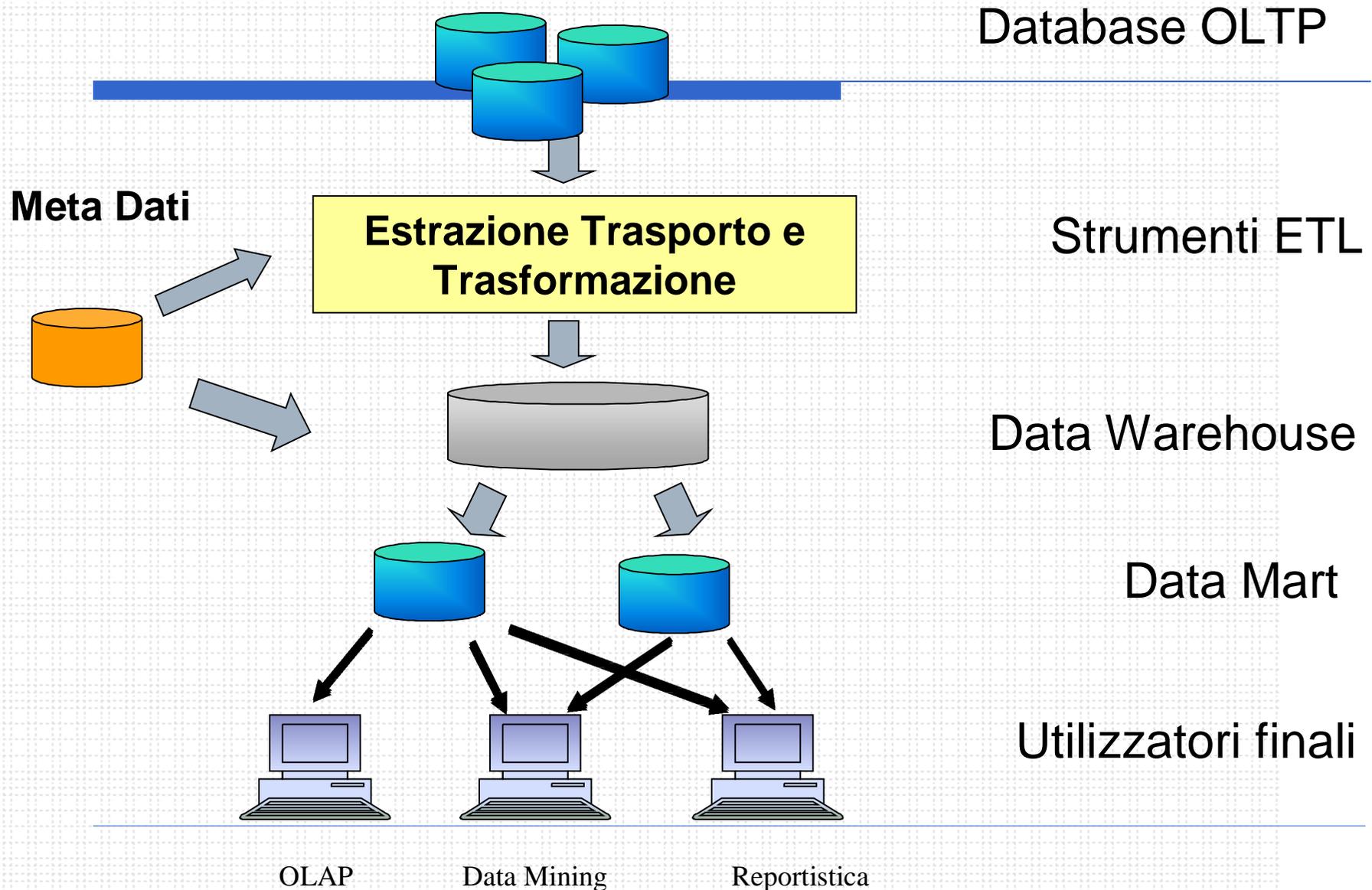
- Il primo punto debole di questa architettura è che non rispetta il requisito di separazione tra l'elaborazione analitica OLAP e quella transazionale OLTP.
- Le interrogazioni di analisi vengono ridirette sui dati operazionali dopo essere state reinterpretate nel middleware, interferendo così con il normale carico di lavoro transazionale

## A due livelli

---

- Nata per evidenziare la separazione tra il livello delle sorgenti e quello del DW, in realtà essa si articola in **quattro** livelli distinti che descrivono stadi successivi del flusso di dati.

## Architettura Multi-Tier (a due livelli)



# 1. Livello sorgenti

---

- Il DW utilizza fonti di dati eterogenei: estratti dall'ambiente di produzione, e quindi originariamente archiviati in database aziendali relazionali oppure provenienti da sistemi informativi esterni all'azienda.

## 2. Livello dell'alimentazione

---

- I dati memorizzati nelle sorgenti devono essere estratti, ripuliti per eliminare le inconsistenze e completare eventuali parti mancanti, integrati per fondere sorgenti eterogenee secondo uno schema comune. I cosiddetti 'strumenti di ETL' (**Extraction-Transformation-Loading**) permettono di integrare schemi eterogenei, nonché di estrarre, trasformare, ripulire, validare, filtrare e caricare i dati dalle sorgenti nel DW.

### 3. Livello del data warehouse

---

- Le informazioni vengono raccolte in un singolo “contenitore” centralizzato logicamente: il DW.
- Esso può essere direttamente consultato ma anche usato come sorgente per costruire **data-mart**, che ne costituiscono una parziale replica, orientati verso specifiche aree dell’impresa.
- Accanto al DW, il contenitore dei **meta-dati** mantiene informazioni sulle sorgenti, sui meccanismi di accesso, sulle procedure di pulizia e alimentazione, sugli utenti, sugli schemi dei data-mart ecc.

## 4. Livello di analisi

---

- Permette la consultazione efficiente e flessibile dei dati integrati a fini di stesura di report, di analisi, di simulazione.

## Definizione di Data-mart

---

- Con il termine Data-mart si intende un sottoinsieme o un'aggregazione dei dati presenti nel DW primario, contenente l'insieme delle informazioni rilevanti per una particolare area del business, una particolare divisione dell'azienda, una particolare categoria di soggetti.
- I data-mart alimentati da DW primario sono spesso detti 'dipendenti'.

# Perché i Data-mart

---

**Sebbene non strettamente necessari essi però costituiscono un'utile risorsa:**

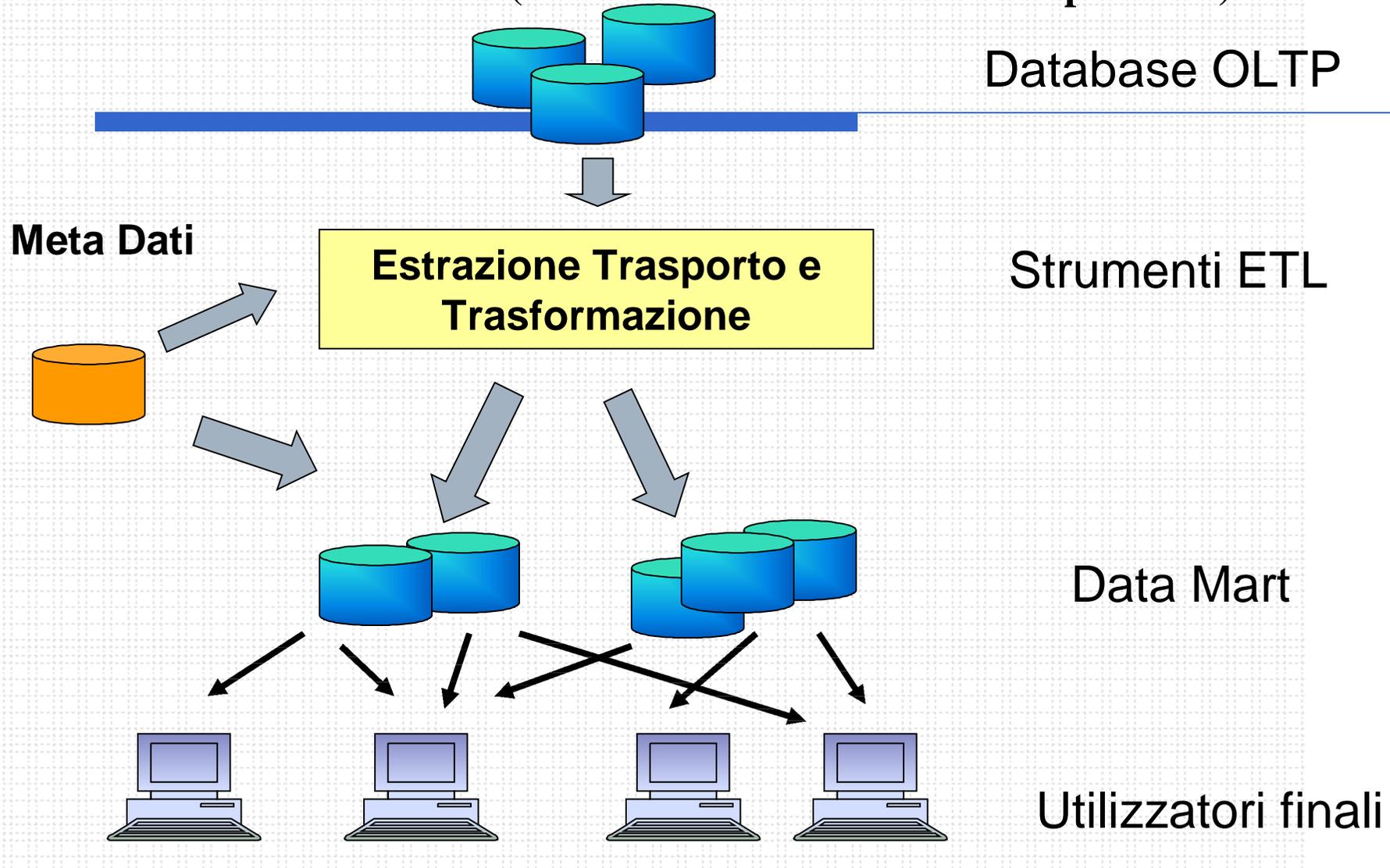
- come blocchi costitutivi durante la realizzazione incrementale del DW,
  - in quanto delineano i contorni delle informazioni necessarie a un particolare tipo di utenti per le loro interrogazioni,
  - poiché, essendo di dimensioni inferiori al DW primario, permettono di raggiungere prestazioni migliori.
-

## A due livelli (con data-mart indipendenti)

---

- In alcuni contesti, per motivi organizzativi e politici, si preferisce adottare un'architettura per il DW in cui i data-mart vengono alimentati direttamente dalle sorgenti e vengono pertanto detti **indipendenti**.

## Architettura Multi-Tier (a due livelli con data mart indipendenti)

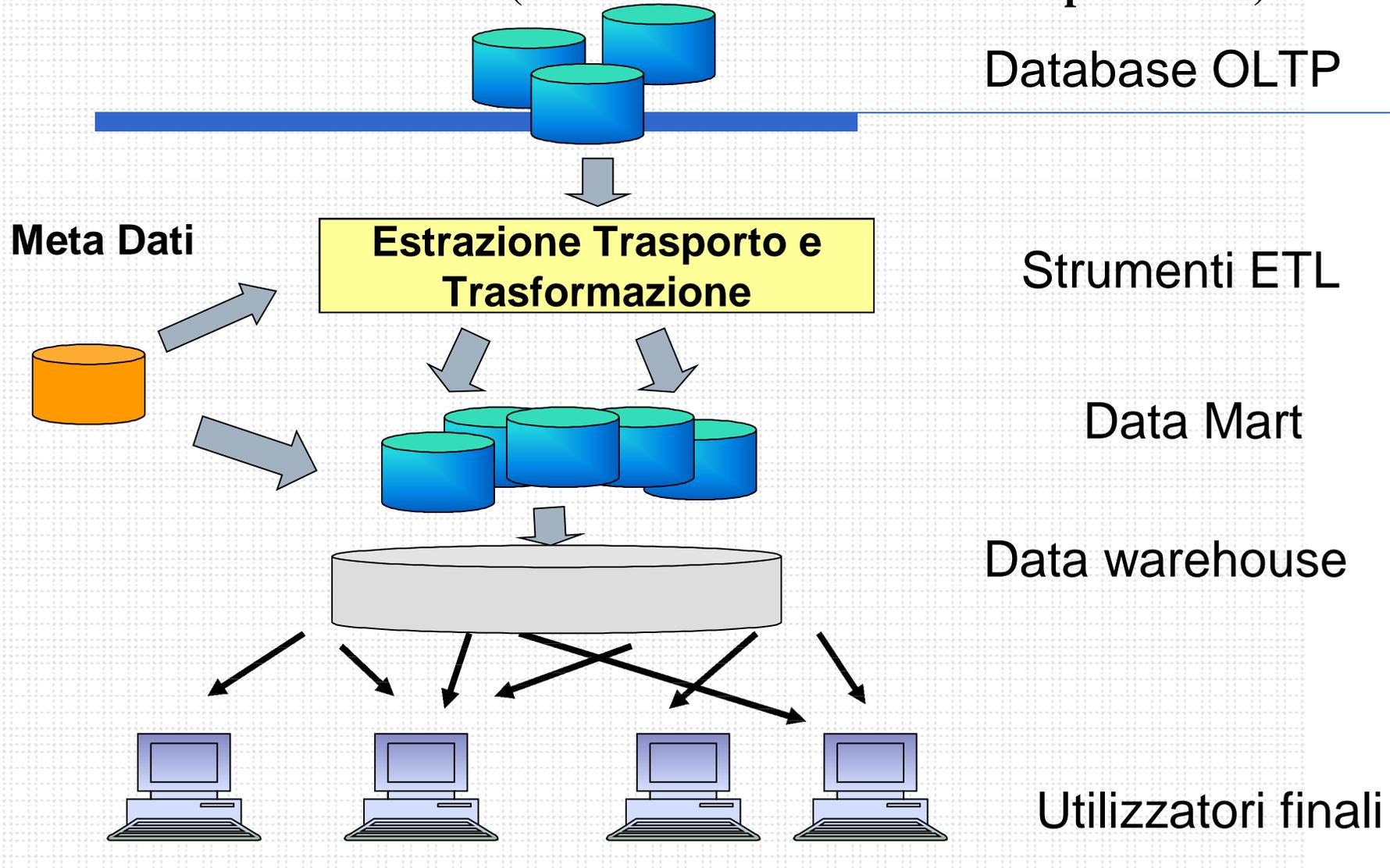


# Problemi

---

- L'assenza di un DW primario snellisce le fasi progettuali, ma determina uno schema complesso di accessi ai dati e ingenera il rischio di inconsistenze tra i data-mart.
  - A volte, pur rispettando l'indipendenza dei data-mart, si preferisce allora creare comunque un DW centrale, ma in questo contesto i ruoli dei data-mart e del DW sono invertiti: infatti il DW viene alimentato dai data-mart e può essere direttamente interrogato al fine di semplificare i pattern degli accessi.
-

## Architettura Multi-Tier (a due livelli con data mart indipendenti-2)

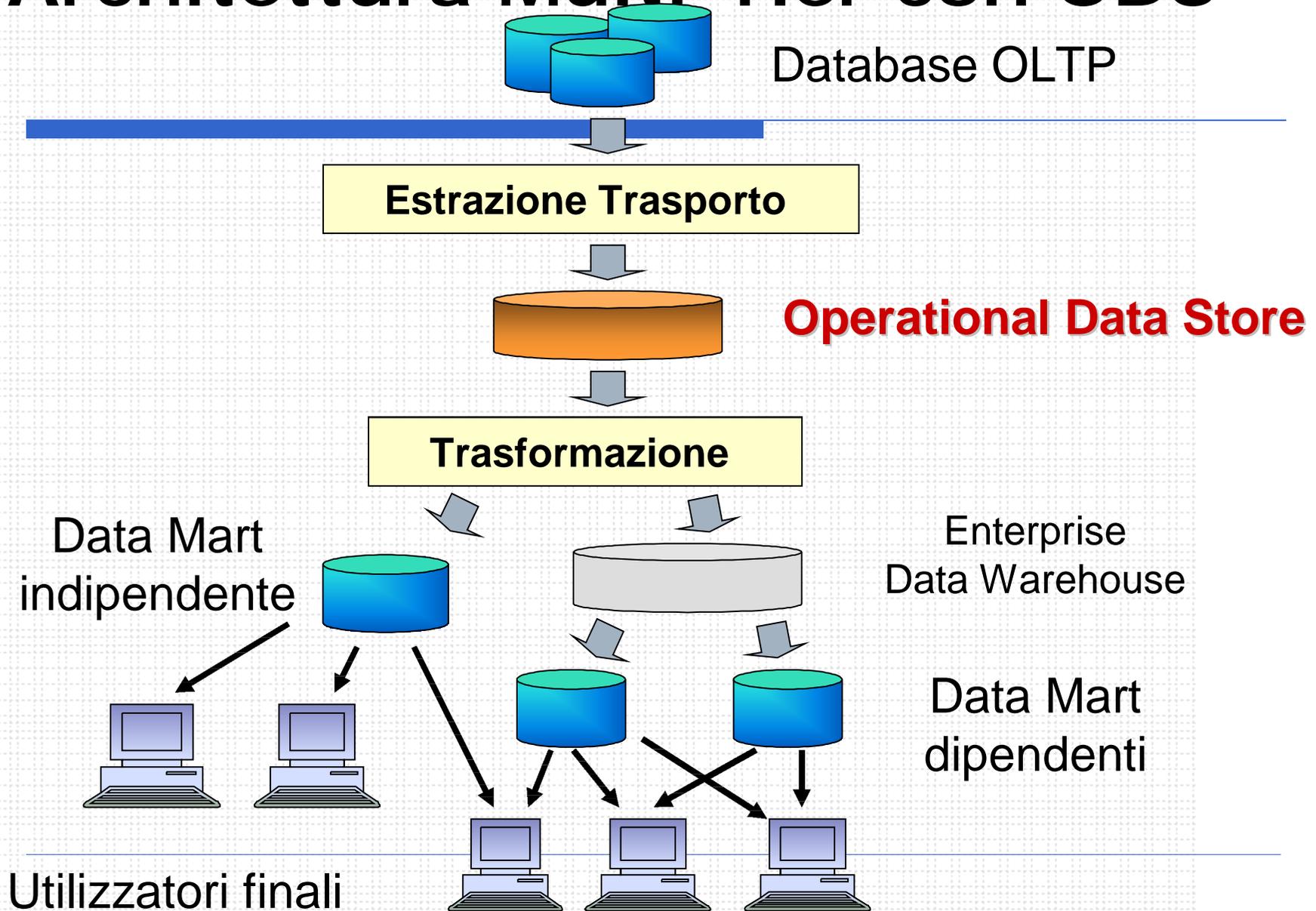


# A tre livelli

---

- Il terzo livello introdotto è quello dei dato riconciliati, detto **Operational Data Store**, che materializza i dati operazionali ottenuti dal processo di integrazione e ripulitura dei dati sorgente: quindi integrati, consistenti, corretti, volatili, correnti e dettagliati.

# Architettura Multi-Tier con ODS



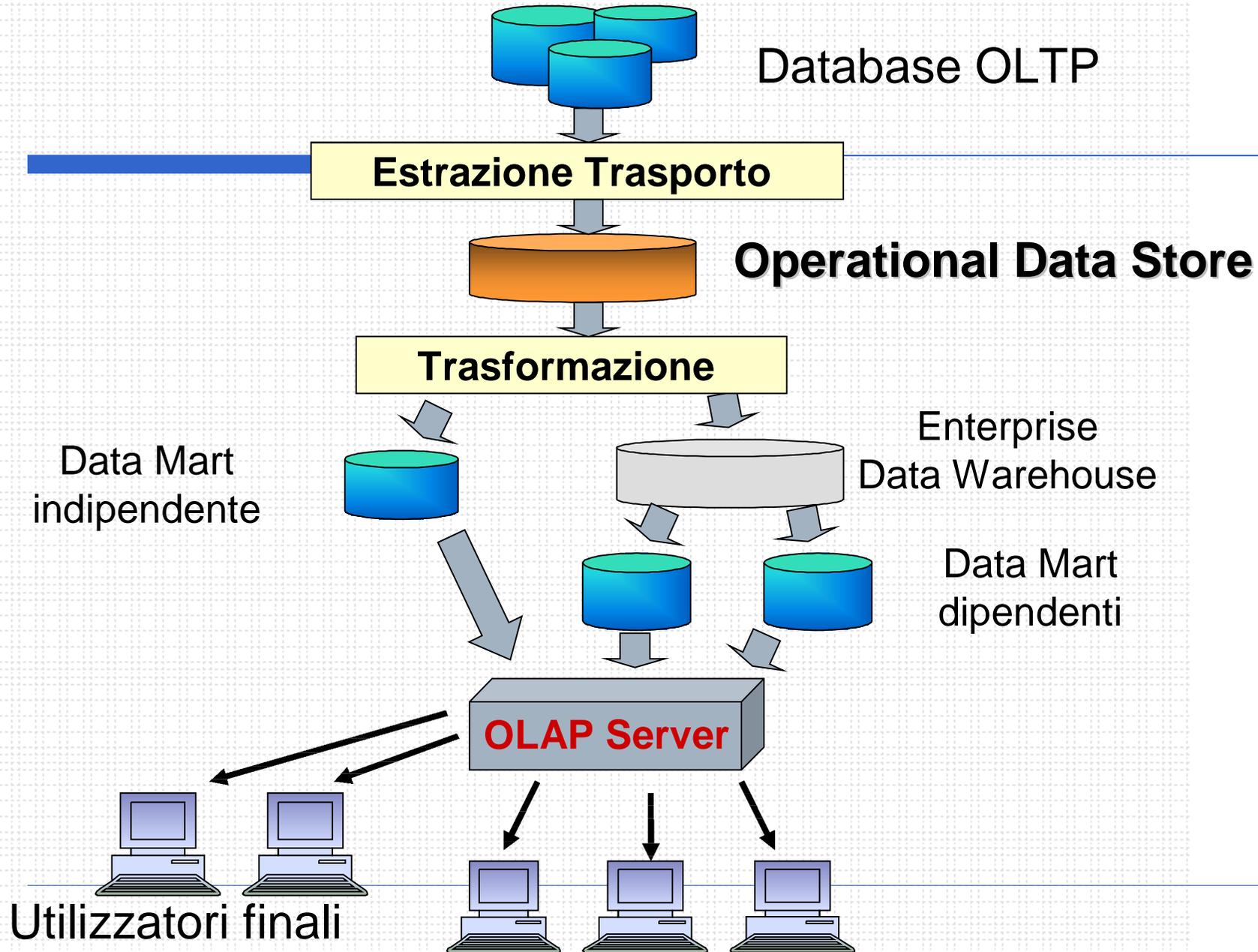
- 
- Il vantaggio principale del livello dei dati riconciliati è che esso crea un modello di dati comune e di riferimento per l'intera azienda, introducendo al contempo una separazione netta tra le problematiche legate all'estrazione e integrazione dei dati dalle sorgenti e quelle inerenti l'alimentazione del DW.
-

# A tre livelli con Olap Server

---

- Infine, nell'architettura più complessa viene introdotto un'ulteriore livello: si tratta dell'utilizzo di un OLAP SERVER che si posiziona tra il livello warehouse (o data-mart) e il livello utente.
- Il cui obiettivo è facilitare le interrogazioni di tipo OLAP sui dati strutturati a livello warehouse.

# Architettura Multi-Tier con ODS e OLAP Server



# Struttura informativa centrata sul DW

