

Processo per la costruzione di informazioni per il supporto decisionale

Sistemi di data warehouse e sistemi di governo

Obiettivi.

Presentare l'evoluzione dei sistemi informativi: da supporto alla operatività a supporto al momento decisionale

Definire il concetto di On Line Analytical Processing (OLAP)

Identificare usi ed utilizzatori di un sistema di Data Warehouse da sistemi settoriali specifici a sistemi integrati.

Illustrare le caratteristiche di un database per il Governo

Presentare i Multidimensional Database ed il modello Data Cube

Definire le modalità di accesso, analisi e reporting: processi e strumenti

Rappresentare l'importanza della Qualità dei dati nei Sistemi di Governo

Fulvio Sbroiavacca



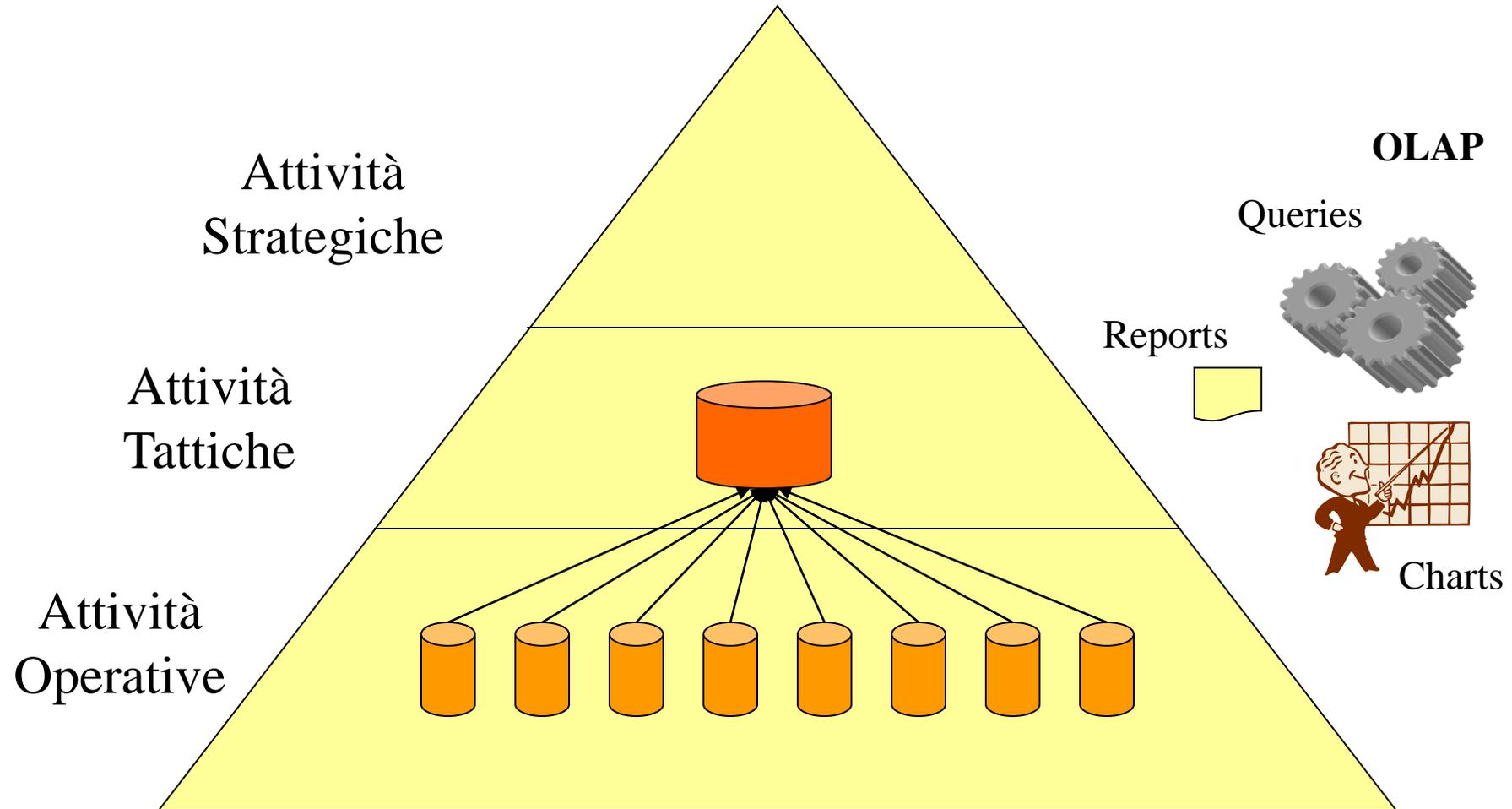
Elaborare dati, produrre informazioni

- Le organizzazioni investono tempo e danaro nei sistemi informativi per le operazioni (operational systems) che gestiscono le informazioni necessarie all'operatività
- Le informazioni sono un **“asset”** dell'organizzazione che deve essere gestito ed utilizzato a fini competitivi
- Per supportare le decisioni strategiche l'organizzazione si dota di sistemi di supporto alle decisioni costruiti su insiemi di informazioni rivolte a tali attività
- Un sistema di supporto alle decisioni integra le informazioni gestionali, i dati storici e statistici attraverso apposite applicazioni, funzioni di accesso ai dati, funzioni di rappresentazione sintetica e restituzione grafica delle informazioni

Sistemi di Data Warehouse

- Sono i sistemi informativi per la raccolta e l'integrazione di dati provenienti da *fonti diverse* con *l'obiettivo di ottenere una visione integrata e dettagliata di un sistema economico*
- La raccolta dei dati ha le caratteristiche di essere **permanente, variabile nel tempo** ed **orientata a specifici argomenti a supporto dell'azione decisionale**

Data Warehouse



OLTP verso OLAP

- *On Line Transaction Processing (OLTP)*:
i sistemi gestionali sono caratterizzati da applicazioni basate su transazioni strutturate, ripetitive, normalmente brevi
- *On Line Analytical Processing (OLAP)*:
i sistemi di supporto decisionale sono caratterizzati da interrogazioni che riguardano grandi moli di dati, spesso complesse

Sistemi gestionali e sistemi di supporto alle decisioni (1)

- Un Data Warehouse mette a disposizione un database che stà alla base dei sistemi di supporto decisionale
- I database gestionali sono disegnati per *automatizzare le attività giornaliere*, sono disegnati per ottimizzare l'input dei dati, in modo sicuro rapido ed efficiente
- Il database di warehouse è disegnato per *aiutare gli utilizzatori a prendere le migliori decisioni attraverso l'analisi delle informazioni* consolidate dell'organizzazione, contiene informazioni riferite ad uno specifico momento temporale, le serie storiche delle informazioni caratterizzanti dell'azienda
- Il data warehouse è finalizzato alle attività di comparazione, analisi e verifica degli andamenti passati e futuri (trends) (ad esempio la comparazione delle vendite tra differenti aree territoriali, oppure la dinamica temporale delle vendite di un prodotto)

Sistemi gestionali e sistemi di supporto alle decisioni (2)

- I dati gestionali sono costantemente in fase di aggiornamento e quindi altamente volatili
- I dati analitici (contenuti nel data warehouse) non sono soggetti a questi cambiamenti (ad esempio il totale dei prodotti venduti alla fine del giorno precedente), permangono e vengono aggiornati secondo procedure predefinite, normalmente vengono accessati in sola lettura (read-only)
- I database gestionali sono progettati per:
 - inserire e memorizzare dati
 - aggiornare dati esistenti
 - controllare transazioni
 - produrre reports
 - mantenere l'integrità dei dati
 - gestire le transazioni
- I database deputati al controllo sono progettati per:
 - grandi volumi di dati in sola lettura per l'utilizzo ai fini decisionali
 - decisioni strategiche (lungo termine) analizzando serie temporali
 - tattiche (breve termine) analizzando e variando "fotografie" (riferite allo stato attuale) di dati quantitativi

Usi ed utilizzatori del Data Warehouse (1)

- Le aree aziendali che possono trarre il maggior benefici dai sistemi di DW sono: finanza, analisi delle vendite, marketing, identificazione dei clienti (customer profiling), analisi di mercato
- Si tratta di aree nelle quali le aziende cercano di impostare gli elementi della propria competitività e quindi dei propri ritorni
- Inizialmente si trattava di database finalizzati al marketing focalizzati su specifiche attività (prodotti più venduti, totale del venduto per aree ecc.), l'evoluzione ha portato poi a sfruttare i database gestionali al fine di trarre le informazioni adeguate per migliorare l'identificazione del "target" di utenza per servizi e prodotti

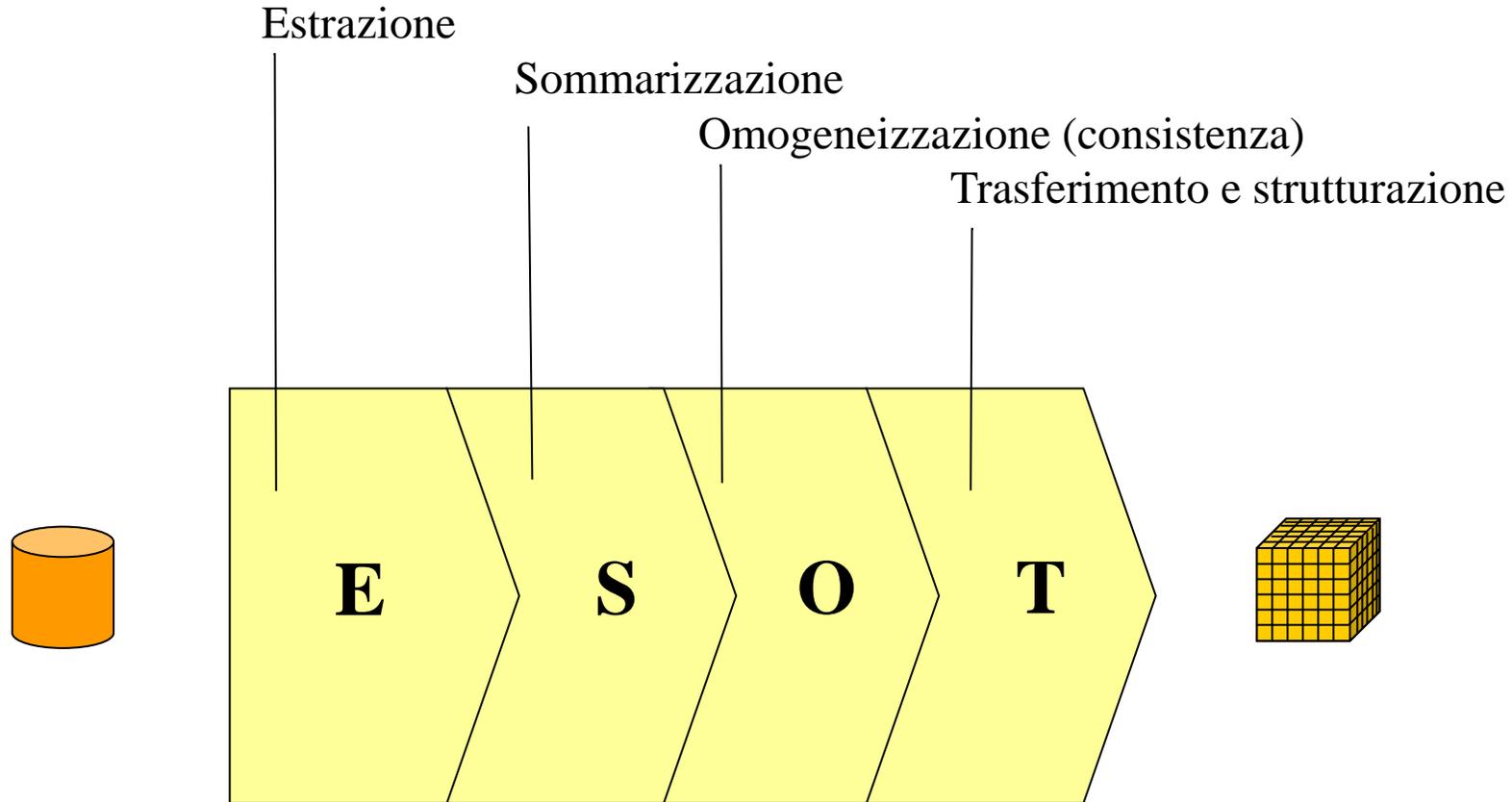
Usi ed utilizzatori del Data Warehouse (2)

- In questa direzione si situano i sistemi dedicati alla “cura” dei clienti denominati Customer Relationship Management (CRM)
- Naturalmente l’evoluzione dei sistemi di DW passa per diverse fasi
 - *dall’iniziale dipendenza da strutture dedicate all’IT delle funzioni di interrogazione messe a disposizione*
 - *alla possibilità di costruire dinamicamente richieste sofisticate e di facile utilizzo (easy to use) da parte degli utilizzatori finali (evoluti nel frattempo verso i cosiddetti knowledge worker)*

Il “processo” di costruzione del Data Warehouse

- Il compito dei sistemi di DW è il consolidamento dei dati e la loro messa a disposizione:
 - i dati *estratti* da vari processi gestionali,
 - devono essere *sommarizzati*,
 - *resi consistenti (omogeneizzati)*,
 - e *trasferiti* in un sistema progettato specificatamente per il supporto decisionale,
 - in modo da consentire l'*utilizzo* di queste informazioni senza conseguenze sulle performance del sistema gestionale (o di produzione)

Il “processo” di costruzione del Data Warehouse

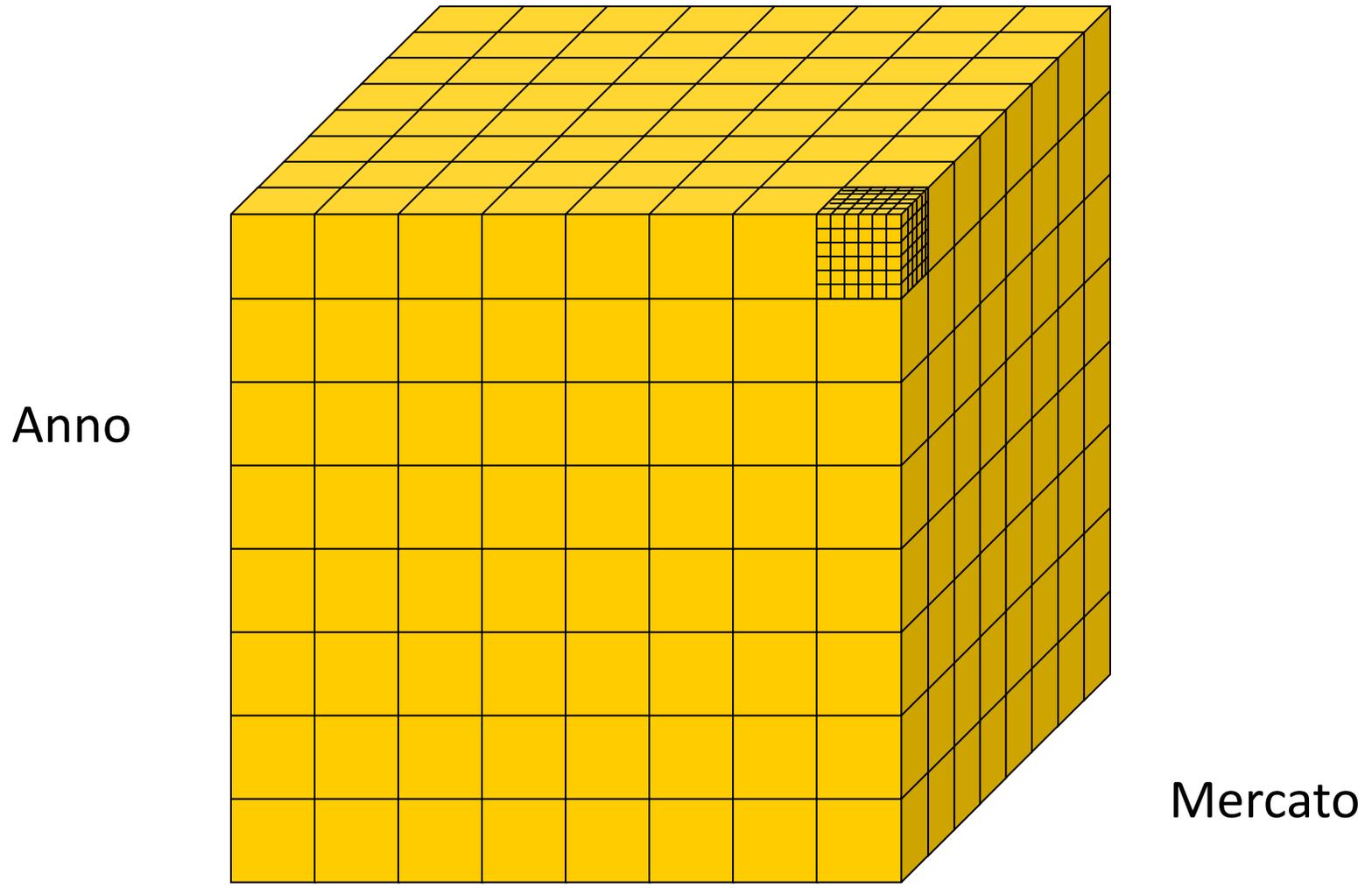


ETL (Extract, Transform, Load) : processo di estrazione, trasformazione e caricamento dei dati in un sistema di sintesi (data warehouse, data mart...)

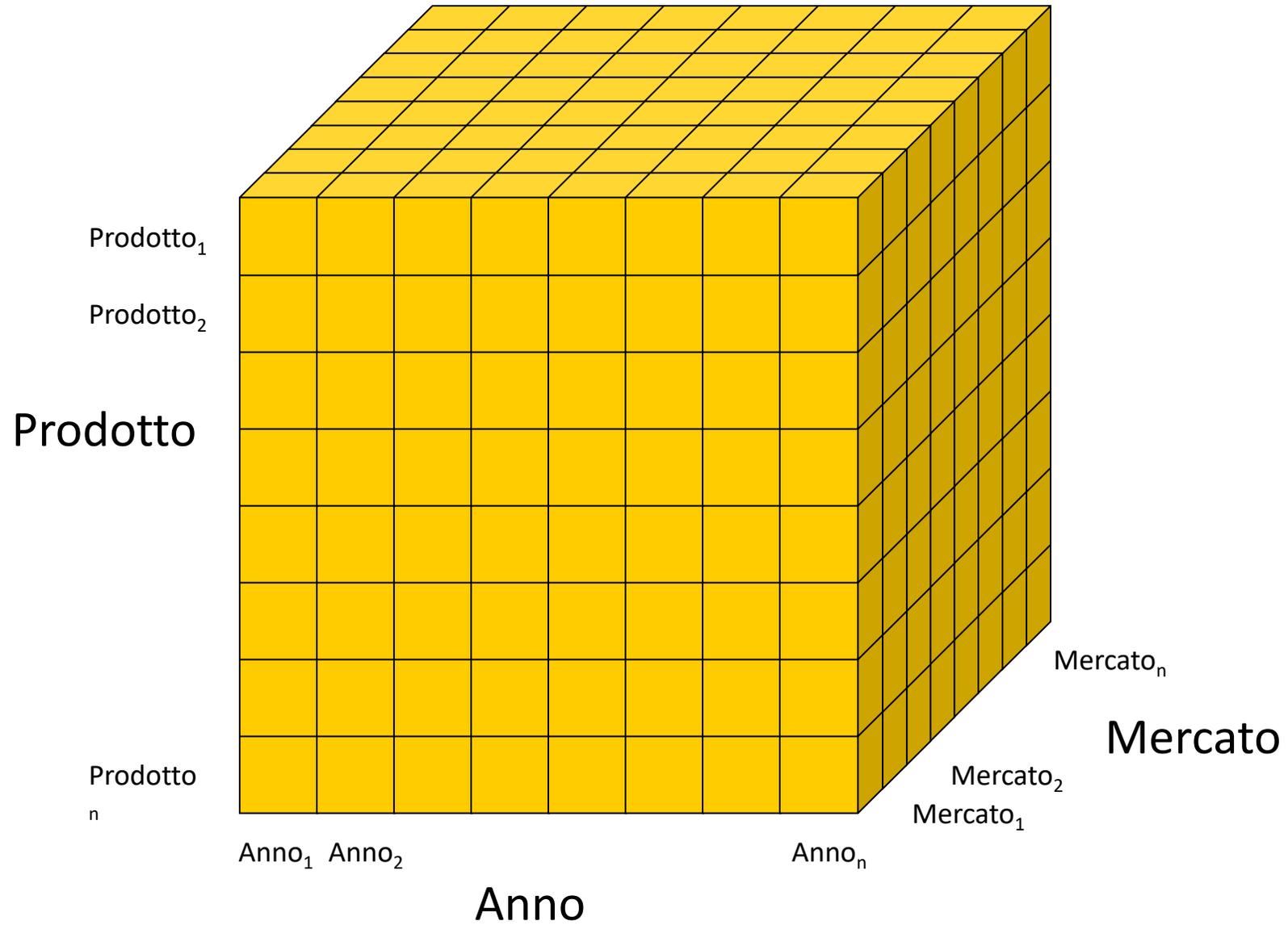
Data CUBE

- Esistono DATABASE progettati appositamente per il data warehouse con funzionalità di strutturazione dei dati che consentono elaborazioni efficienti di grandi moli di dati
- Una tecnica consiste nell'associare ad un database relazionale un database multidimensionale (MDD Multidimensional database)
- I modelli dati adatti alle applicazioni OLAP devono supportare:
 - *analisi*
 - *calcoli sofisticati*
 - *su diverse gerarchie e dimensioni*
- Il modello più adatto è un database multidimensionale: il Data Cube
- L'accesso ad ogni elemento del data cube è molto **efficiente** essendo **costante la dimensione del cammino** che lo identifica

Data Cube

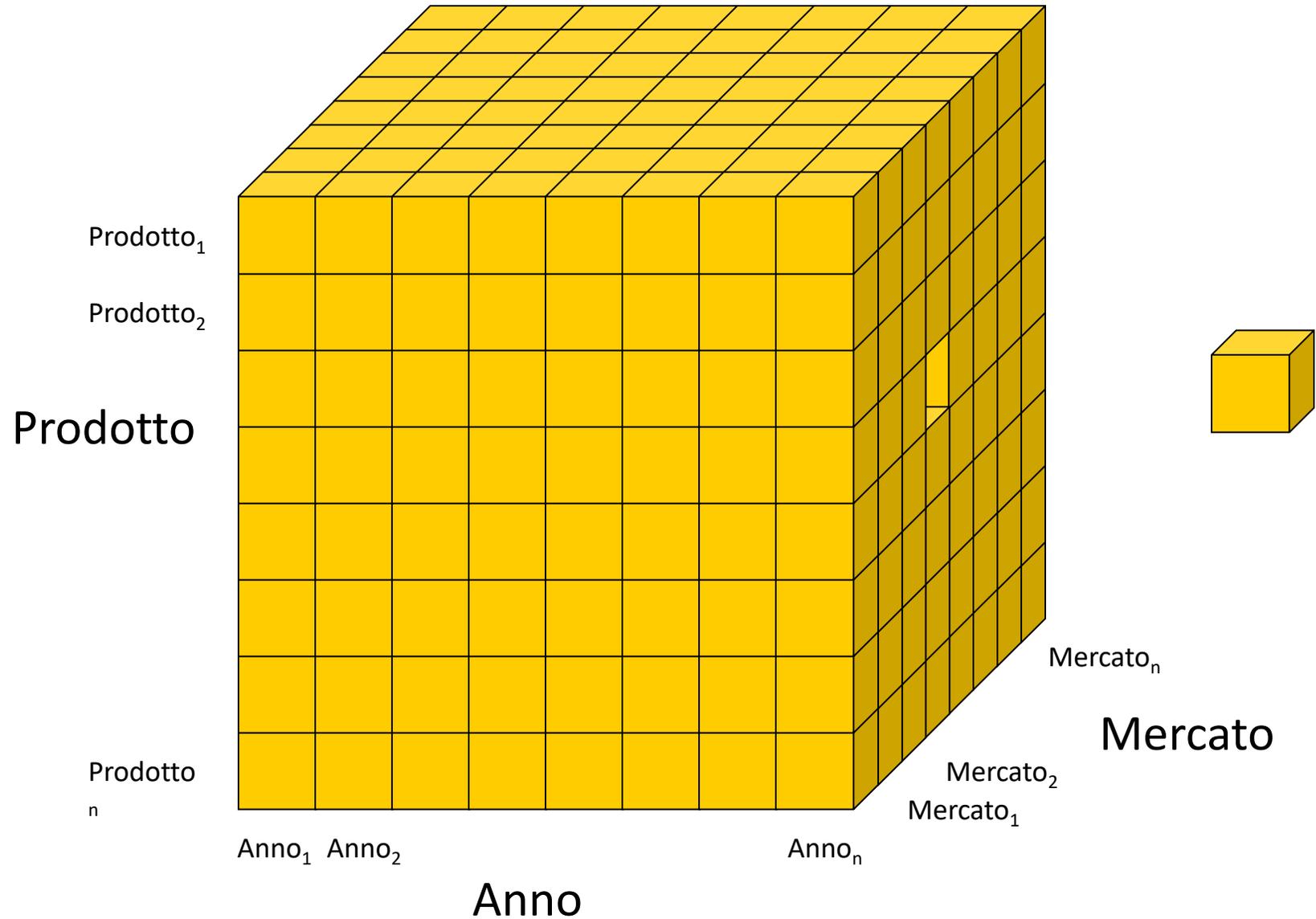


Data Cube

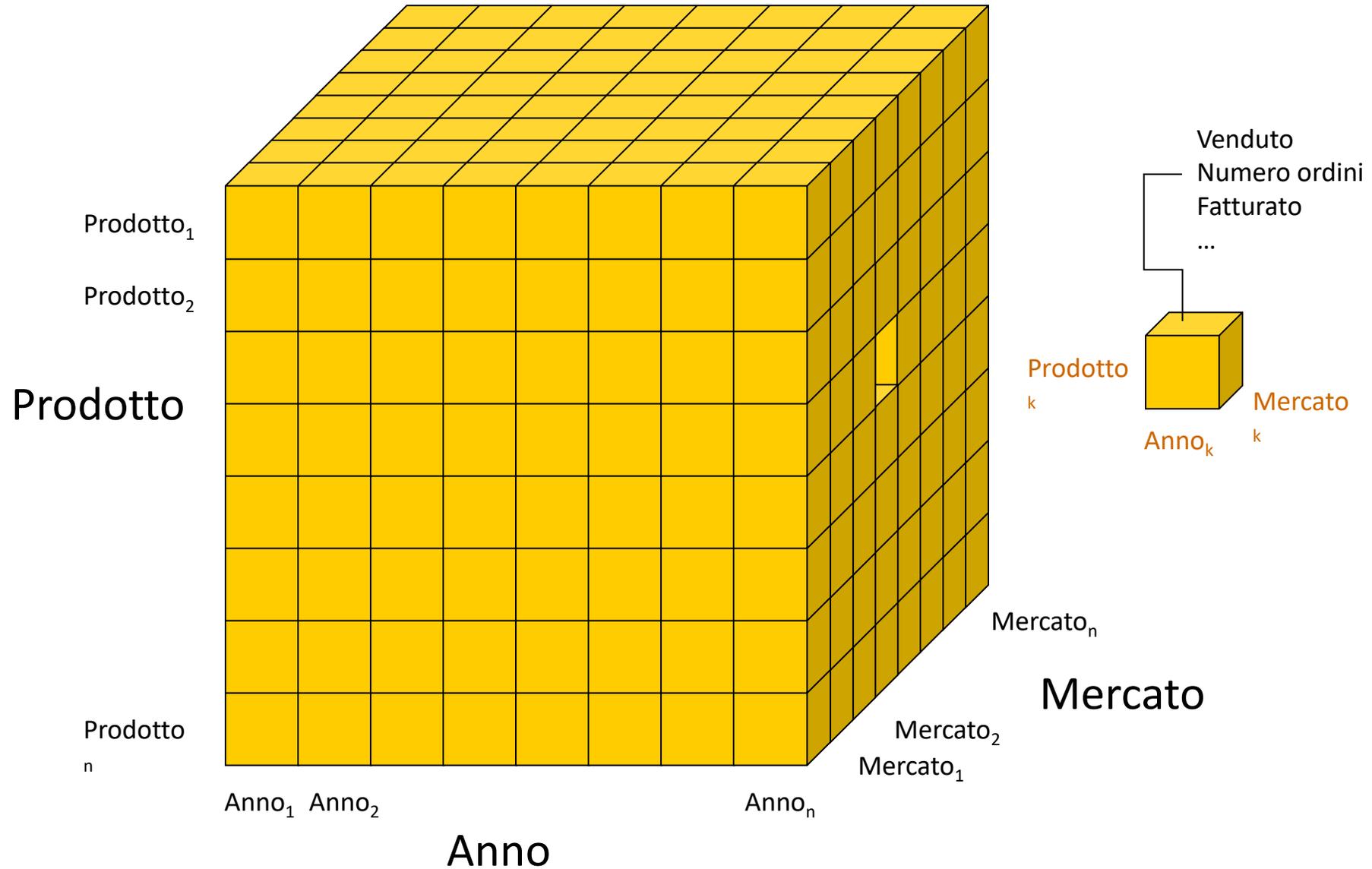


Data Cube

Le celle del cubo contengono i valori da analizzare relativi alle dimensioni che le individuano



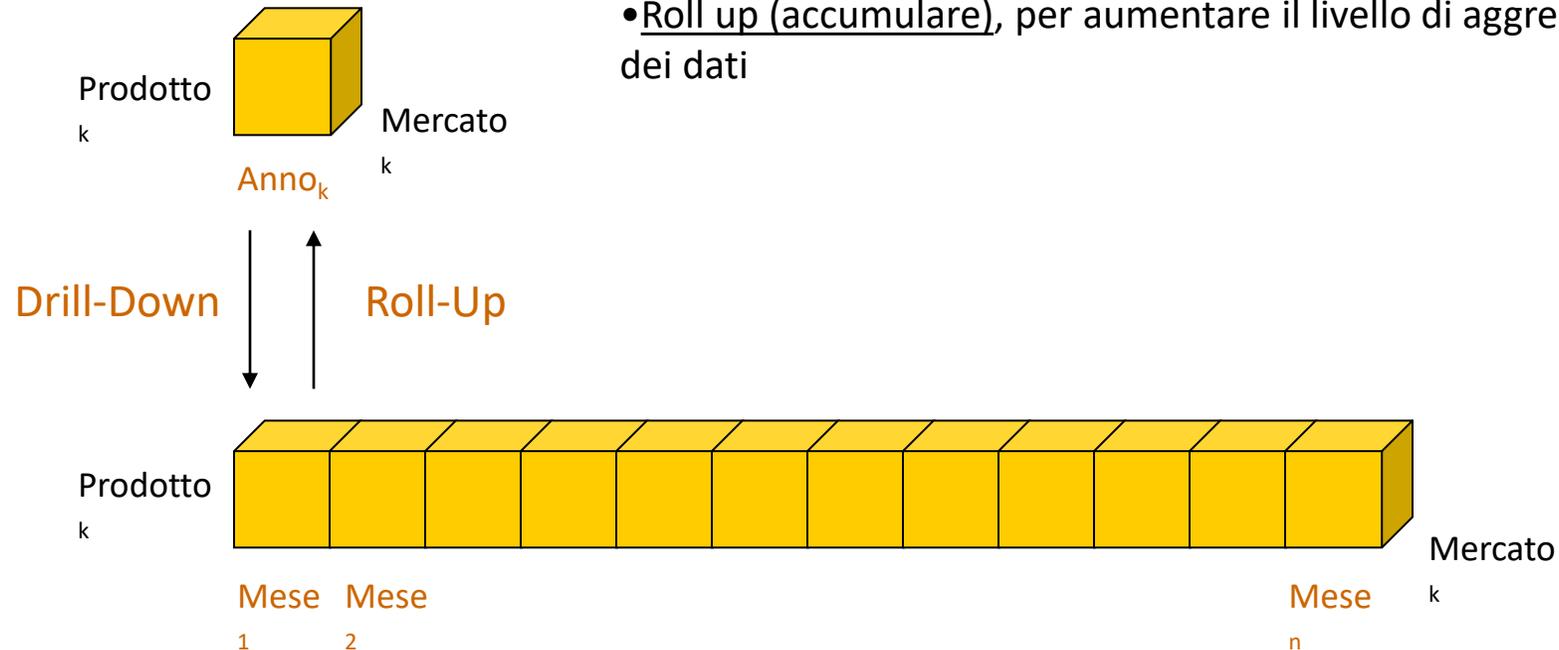
Data Cube

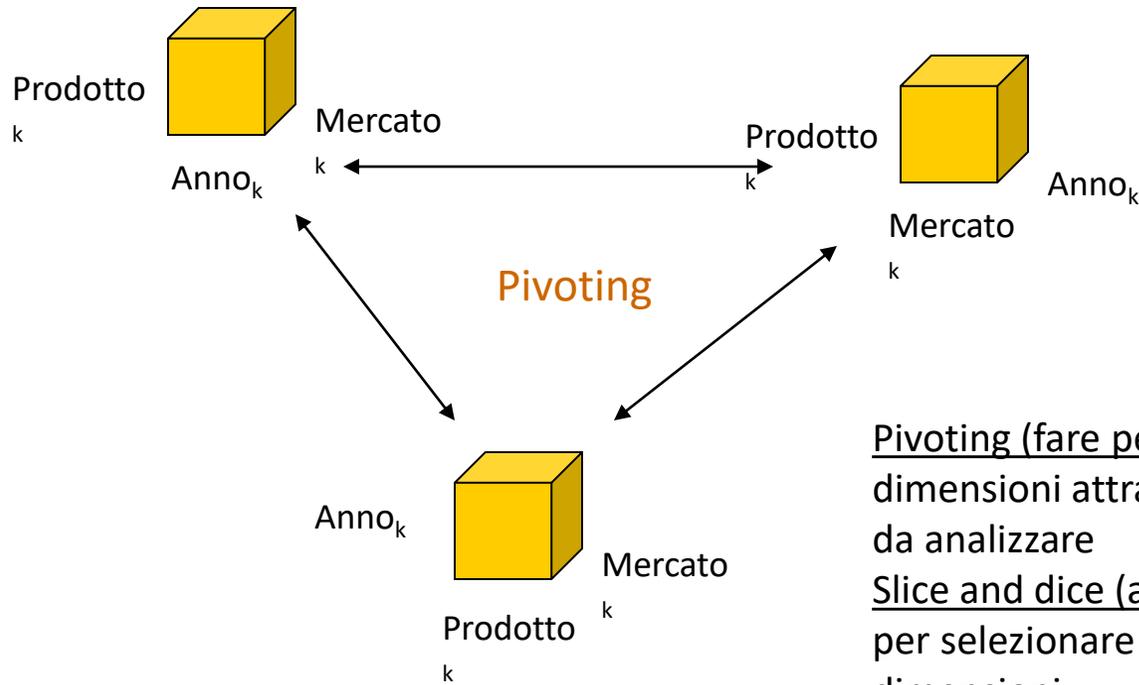


Operazioni sui Data Cube

Le tipiche operazioni (OLAP) che si eseguono sul Data Cube sono:

- Drill down (perforare), per aumentare il livello di dettaglio dei dati
- Roll up (accumulare), per aumentare il livello di aggregazione dei dati





Pivoting (fare perno), per selezionare due dimensioni attraverso le quali aggregare i valori da analizzare

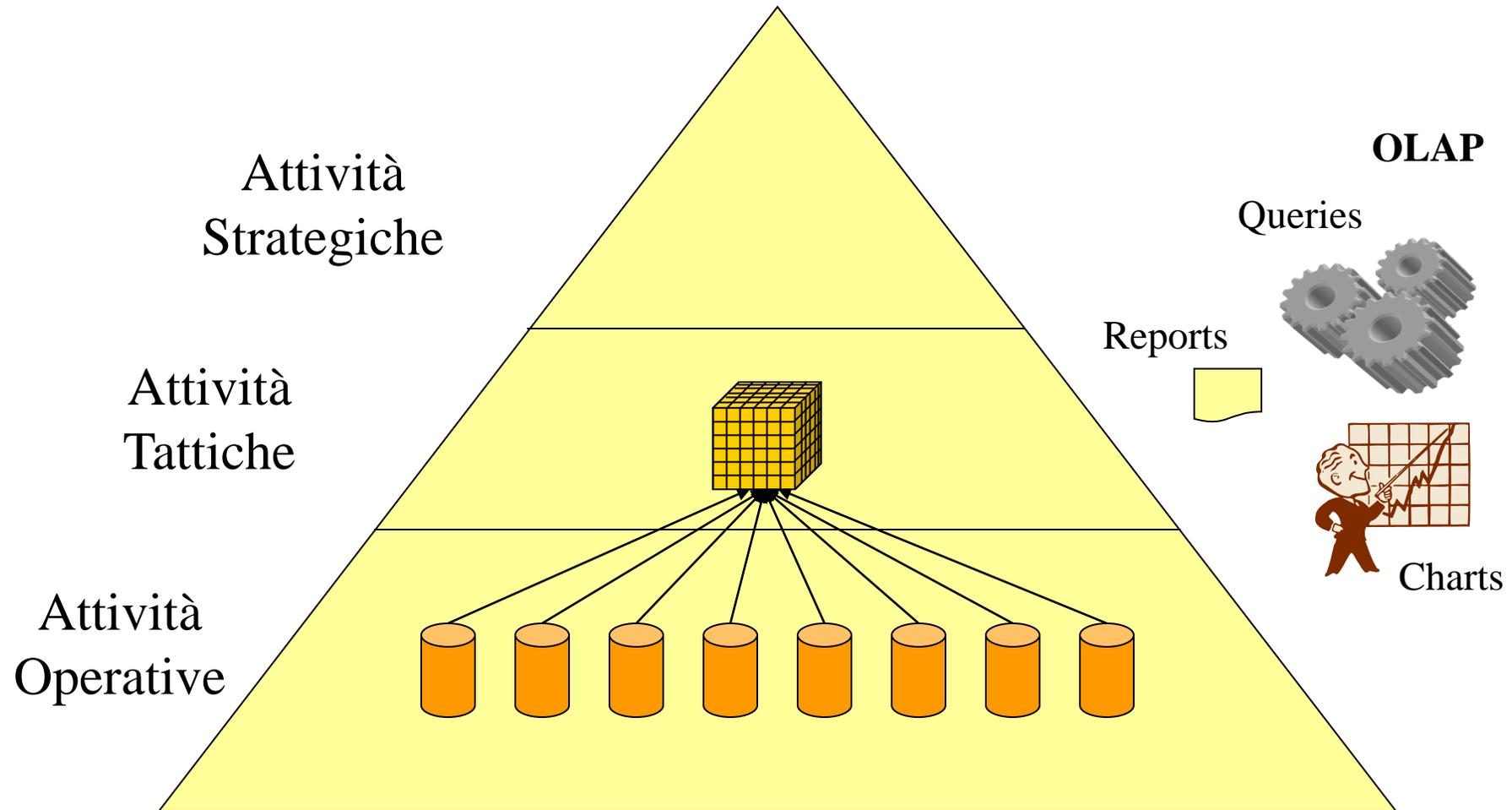
Slice and dice (affettare e ritagliare a cubetti), per selezionare e proiettare i dati riducendo le dimensioni

Ranking (attribuire una classe di merito), per ordinare i dati secondo diversi criteri

Accesso, Analisi e Reporting

- Sono disponibili diverse tipologie di strumenti
 - **Strumenti di front-end** come ad esempio BusinessObjects
 - **Spreadsheets** come ad esempio Microsoft Excel
 - **Sistemi completi** di gestione ed analisi come ad esempio SAS
- Con questi strumenti l'utente finale (end user) può accedere al DW senza conoscere la struttura dei dati o la programmazione SQL
- Può eseguire analisi approfondite sui dati (funzioni di drill down) e copiare i risultati in un spreadsheet per ulteriori analisi

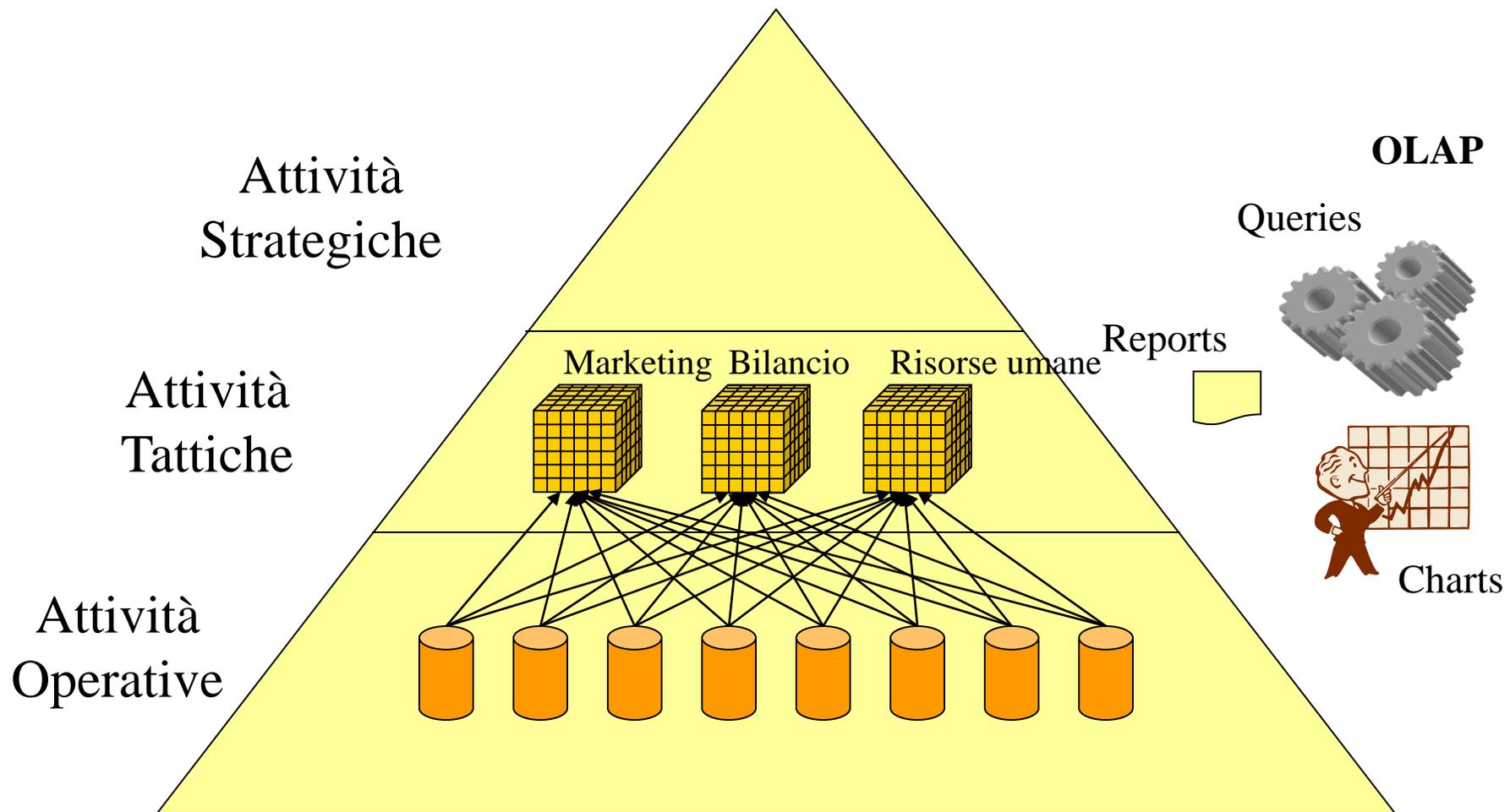
Sistemi di Governo



I Data Marts

- I sistemi di data warehouse comportano costi iniziali (di hardware e di software) che rappresentano solo una parte del costo complessivo
- Bisogna infatti considerare che la produzione di informazioni determina poi l'aumento della richiesta informativa da parte degli utilizzatori
- Il tutto comporta naturalmente significativi **costi** nel ciclo di vita del sistema in particolare per quanto riguarda evoluzione e mantenimento del sistema
- I Data marts rappresentano un'alternativa al data warehouse, si possono considerare come piccoli data warehouse o sottoinsiemi dello stesso
- *Invece di avere un data warehouse centrale, l'azienda si dota di più data warehouse di limitate dimensioni finalizzate ai diversi utilizzatori (data marts)*

Data Marts



Una raccomandazione: **la qualità dei dati**

- I Sistemi di Governo richiedono la disponibilità di dati gestionali di alta qualità, se i dati sono poco affidabili gli utilizzatori saranno portati ad interpretazioni sbagliate
- Per ottenere una buona qualità dei dati è necessario un processo di reingegnerizzazione dei dati che inizia sempre con un'analisi del livello di bontà delle informazioni disponibili e con la specificazione del livello di qualità che si vuole raggiungere
- La reingegnerizzazione dei dati consiste in un'analisi dei dati gestionali, la trasformazione di questi in informazioni integrate e la generazione di database relazionali
- *La qualità delle informazioni è un fattore fondamentale che va valutato prima di avviare lo sforzo per costruire il data warehouse*
- Naturalmente deve essere sempre impostato un processo di monitoraggio e mantenimento della qualità dei dati per assicurare l'efficacia complessiva del Sistema di Governo nel tempo

Quest'opera è stata rilasciata con licenza Creative Commons Attribuzione - Non commerciale - Condividi allo stesso modo 4.0 Internazionale.
Per leggere una copia della licenza visita il sito web <http://creativecommons.org/licenses/by-nc-sa/4.0/>.