



Il Data Warehousing

Prof. Stefano Rizzi

Alma Mater Studiorum - Università di Bologna



BUSINESS INTELLIGENCE

La Business Intelligence (BI) combina

- business analytics,
- data mining,
- visualizzazione dei dati,
- strumenti e infrastrutture per i dati,

nonché le best practice per permettere alle organizzazioni di prendere più decisioni basate sui dati.

Si parla anche di Piattaforma di Business Intelligence, ossia l'insieme delle applicazioni usate dai manager aziendali per

- analizzare lo stato dell'azienda
- prendere decisioni rapide
- prendere decisioni migliori

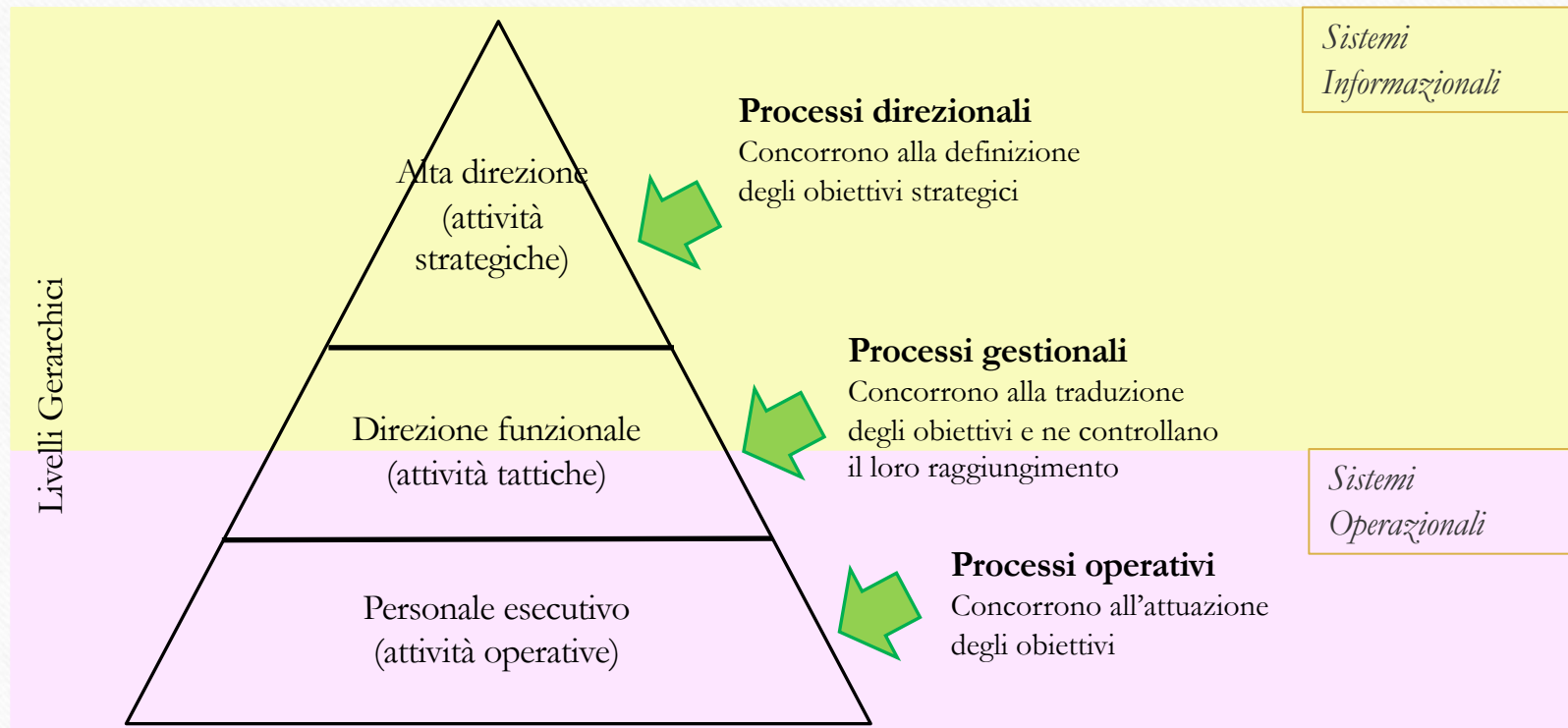
Quindi la Business Intelligence è una disciplina che consente a chi deve decidere in azienda di capire, attraverso soluzioni software, i fattori chiave del business e conseguentemente di prendere le migliori decisioni in quel momento.



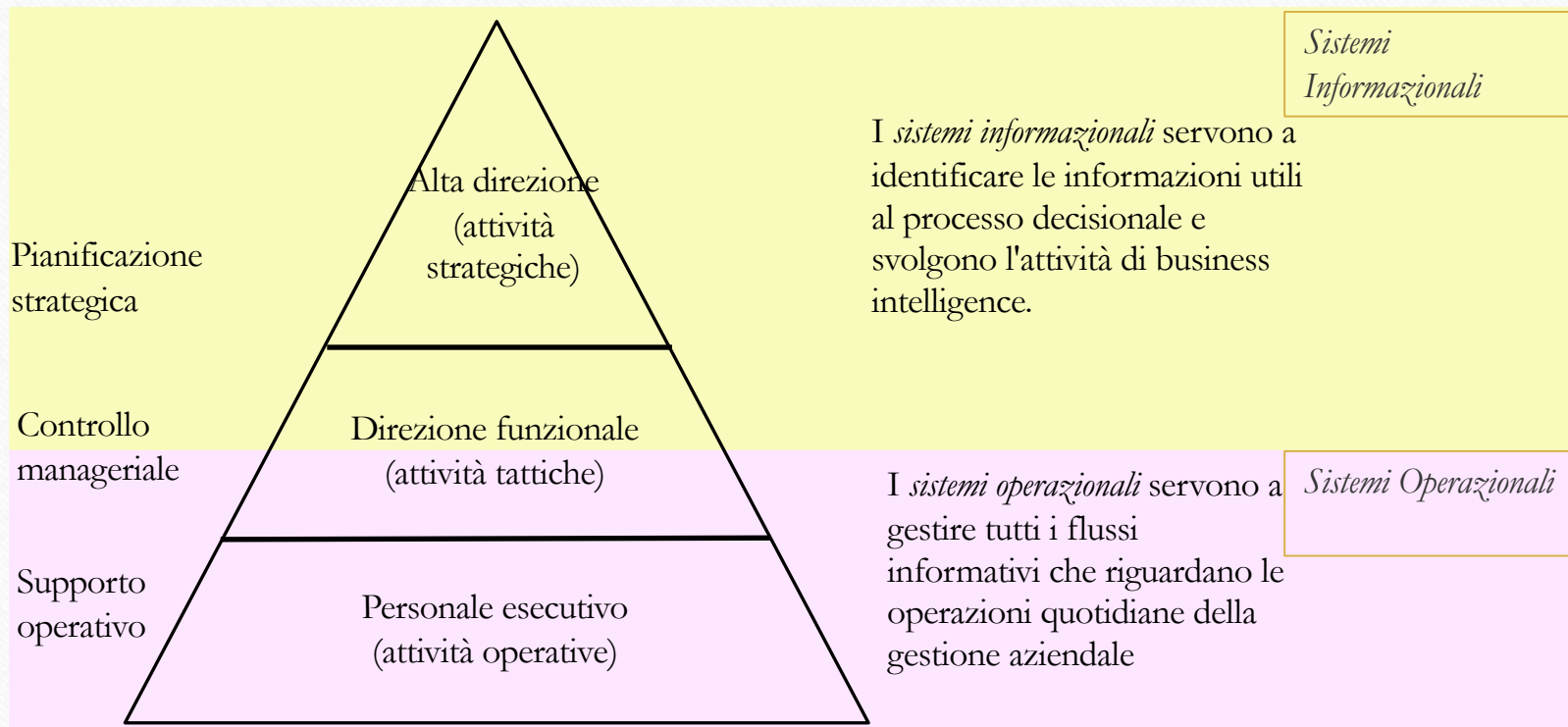
Business intelligence

- Si parla di piattaforma poiché per consentire ai manager analisi potenti e flessibili è necessario definire un'apposita infrastruttura hardware e software di supporto composta da:
 - ✓ Hardware dedicato
 - ✓ Infrastrutture di rete
 - ✓ DBMS
 - ✓ Software di back-end
 - ✓ Software di front-end
- Il ruolo chiave di una piattaforma di business intelligence è la *trasformazione dei dati aziendali in informazioni* fruibili a diversi livelli di dettaglio

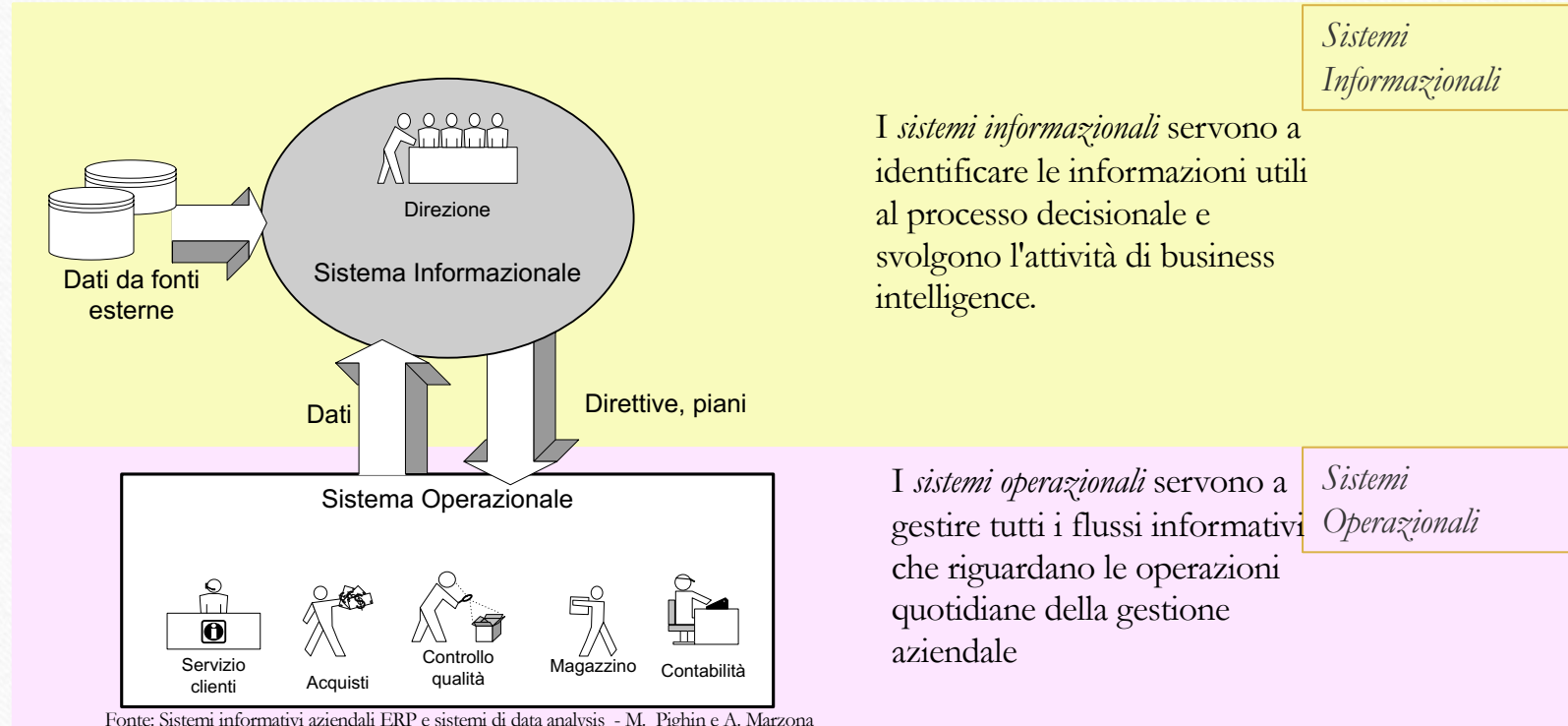
Sistemi Informazionali e Operazionali



Sistemi Informazionali e Operazionali



Sistemi Informazionali e Operazionali



Fonte: Sistemi informativi aziendali ERP e sistemi di data analysis - M. Pighin e A. Marzona



Dai dati alle informazioni

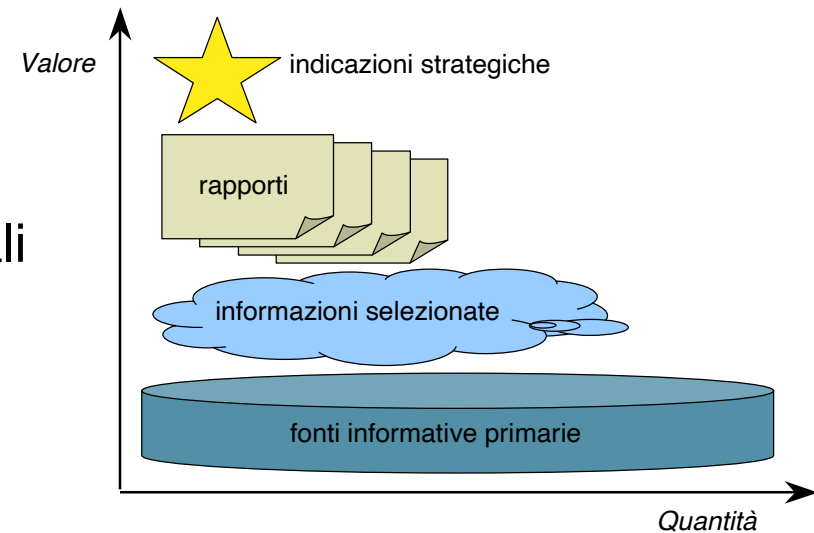
- L'informazione è un bene a valore crescente, necessario per pianificare e controllare le attività aziendali con efficacia
- Essa costituisce la materia prima che viene trasformata dai sistemi informativi, come i semilavorati vengono trasformati dai sistemi di produzione

~~dati = informazione~~

- Spesso la disponibilità di troppi dati rende arduo, se non impossibile, estrapolare le informazioni veramente importanti

Dai dati alle informazioni

- Per ogni azienda è fondamentale poter disporre in maniera rapida e completa delle informazioni necessarie al processo decisionale: le indicazioni strategiche sono estrapolate principalmente dalla mole dei dati operazionali contenuti nei database aziendali, attraverso un procedimento di selezione e sintesi progressiva

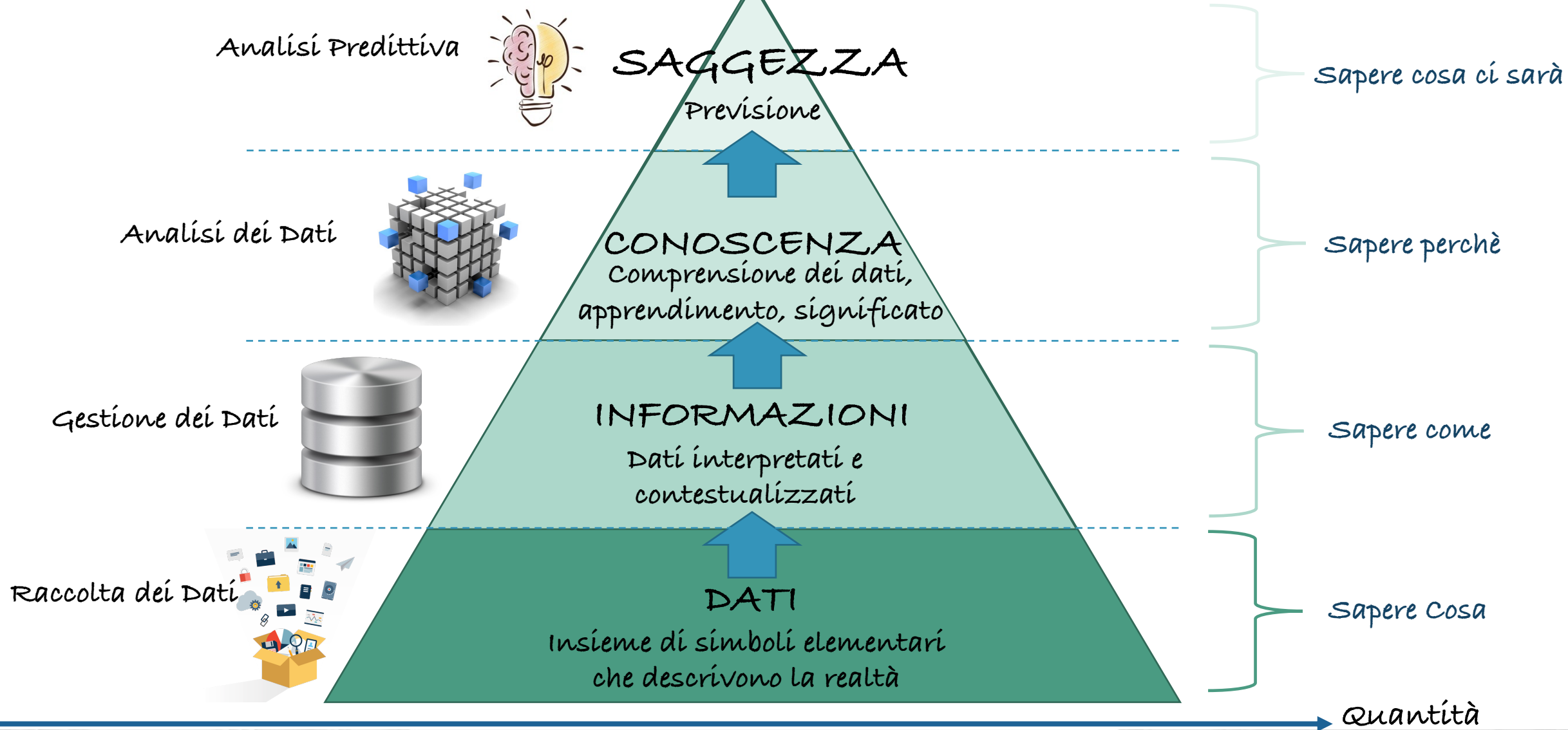




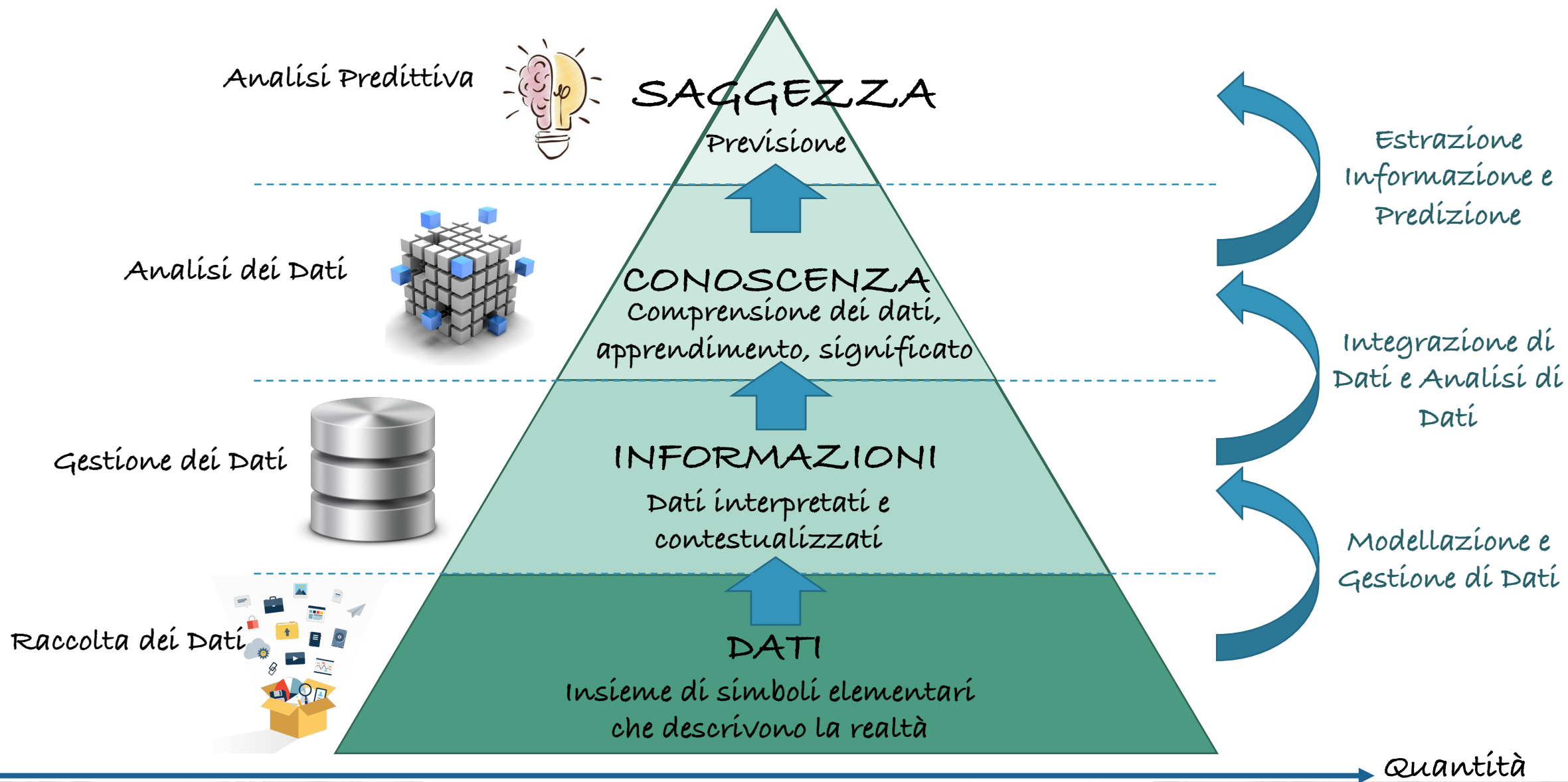
Il nuovo ruolo dell'informatica

- L'aumento esponenziale del volume dei dati operazionali ha reso il calcolatore l'unico supporto adatto al processo decisionale
- L'utilizzo massiccio di tecniche di analisi dei dati aziendali ha reso il sistema informativo un elemento strategico per la realizzazione del business
- Il ruolo dell'informatica è passato da passivo strumento per la registrazione delle operazioni a fattore decisivo per la individuazione di elementi critici dell'organizzazione e di potenziali aree di business

La Piramide dell'Informazione



valore



UNIVERSITÀ
di VERONA

Dipartimento
di INFORMATICA

I sistemi di supporto alle decisioni

- Negli anni '80 nascono i sistemi di supporto alle decisioni (*decision support system*):

l'insieme delle tecniche e degli strumenti informatici atti a estrapolare informazioni da un insieme di dati memorizzati su supporti elettronici

Ruolo del DSS

Nel Passato

Descrivere il passato

Ridurre i costi

Descrivere i problemi

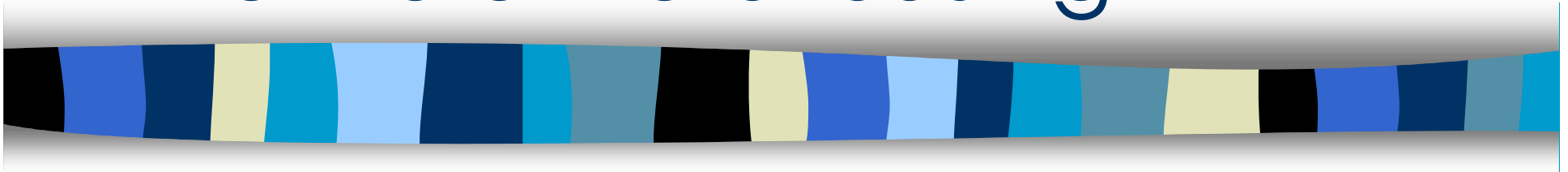
Nel Futuro

Anticipare il futuro

Aumentare i profitti

Suggerire i cambiamenti da apportare

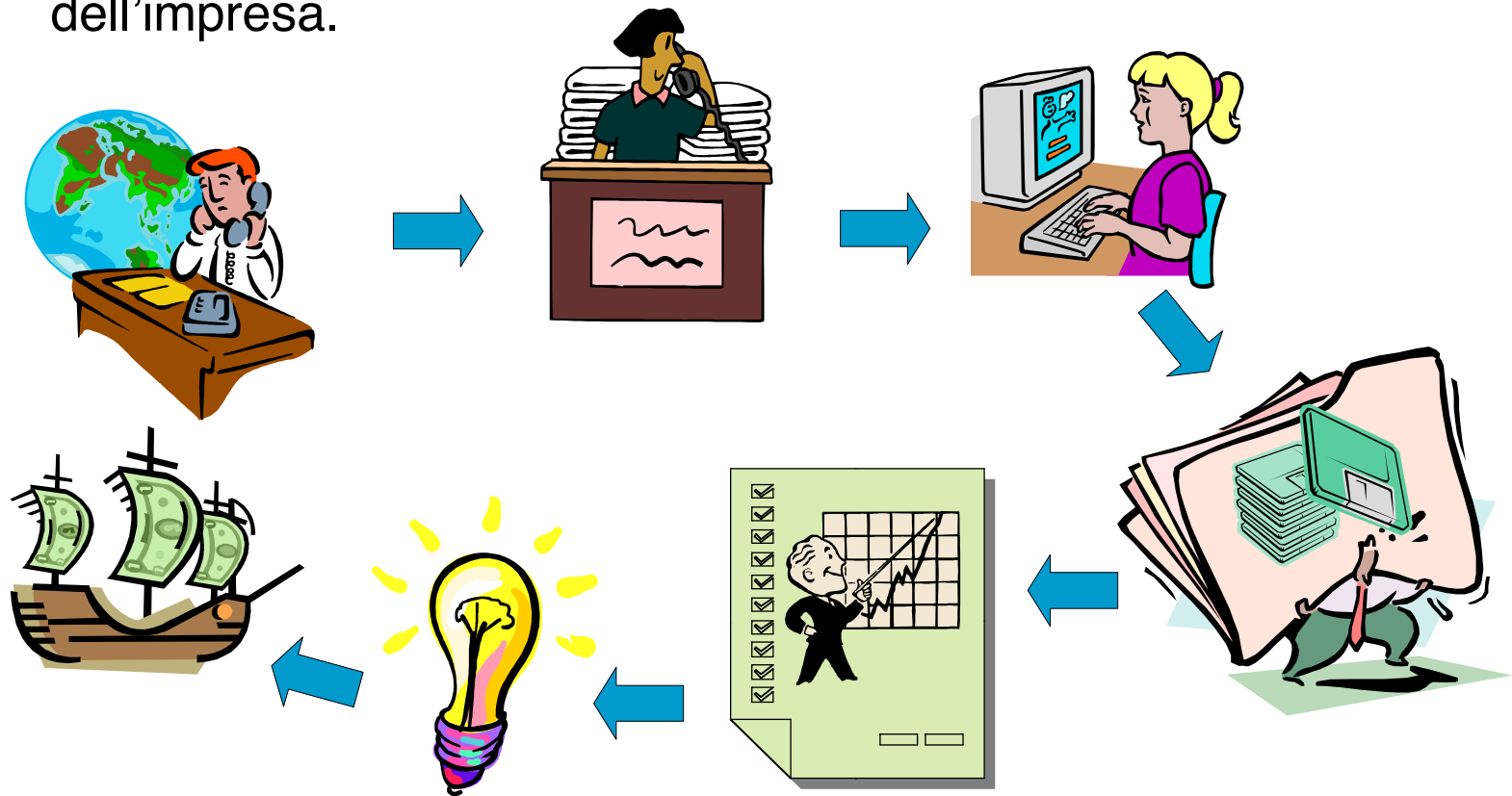
Introduzione al Data Warehousing



Prof. Stefano Rizzi

Uno scenario tipico...

- .. è quello di una grande azienda, con numerose filiali, i cui dirigenti desiderano quantificare e valutare il contributo dato da ciascuna di esse al rendimento commerciale globale dell'impresa.

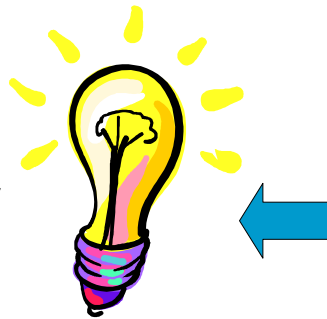


Uno scenario tipico...

- .. è quello di una grande azienda, con numerose filiali, i cui dirigenti desiderano quantificare e valutare il contributo dato da ciascuna di esse al rendimento commerciale globale dell'impresa.



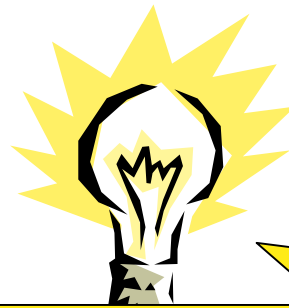
un raccoglitore di informazioni che integra e riorganizza i dati provenienti da sorgenti di varia natura e li rende disponibili per analisi e valutazioni finalizzate alla pianificazione e al processo decisionale



**DATA
WAREHOUSE**

OLTP e OLAP

- Mescolare interrogazioni “analitiche” e “transazionali” di routine porta a inevitabili rallentamenti che rendono insoddisfatti gli utenti di entrambe le categorie.



separare l'elaborazione di tipo analitico (**OLAP, On-Line Analytical Processing**) da quella legata alle transazioni (**OLTP, On-Line Transactional Processing**), costruendo un nuovo raccoglitore di informazioni che integri i dati provenienti da sorgenti di varia natura, li organizzi in una forma appropriata e li renda disponibili per scopi di analisi e valutazione finalizzate alla pianificazione e al processo decisionale



Alcune aree di utilità

- **Commercio** (analisi delle vendite e dei reclami, controllo di spedizioni e inventari, cura del rapporto con i clienti)
- **Manifattura** (controllo dei costi di produzione, supporto fornitori e ordini)
- **Servizi finanziari** (analisi del rischio e delle carte di credito, rivelazione di frodi)
- **Trasporti** (gestione parco mezzi)
- **Telecomunicazioni** (analisi del flusso delle chiamate e del profilo dei clienti)
- **Sanità** (analisi di ricoveri e dimissioni, contabilità per centri di costo)
-



Data Warehousing:

- Una collezione di metodi, tecnologie e strumenti di ausilio al *knowledge worker* (dirigente, amministratore, gestore, analista) per condurre analisi dei dati finalizzate all'attuazione di processi decisionali e al miglioramento del patrimonio informativo.

Le lamentele

- ⇒ *abbiamo montagne di dati ma non possiamo accedervi!*
- ⇒ *come è possibile che persone che svolgono lo stesso ruolo presentino risultati sostanzialmente diversi?*
- ⇒ *vogliamo selezionare, raggruppare e manipolare i dati in ogni modo possibile!*
- ⇒ *mostratemi solo ciò che è importante!*
- ⇒ *tutti sanno che alcuni dati non sono corretti!*

R. Kimball, The Data Warehouse Toolkit





Caratteristiche del processo di warehousing

- **accessibilità** a utenti con conoscenze limitate di informatica e strutture dati;
- **integrazione dei dati** sulla base di un modello standard dell'impresa;
- **flessibilità di interrogazione** per trarre il massimo vantaggio dal patrimonio informativo esistente;
- **sintesi** per permettere analisi mirate ed efficaci;
- **rappresentazione multidimensionale** per offrire all'utente una visione intuitiva ed efficacemente manipolabile delle informazioni;
- **correttezza e completezza** dei dati integrati.



Il Data Warehouse

- Al centro del processo, il data warehouse è un contenitore di dati che si fa garante dei requisiti esposti.

➤ *Un **Data Warehouse** è una collezione di dati di supporto per il processo decisionale che presenta le seguenti caratteristiche:*

- ✓ *è orientata ai soggetti di interesse;*
- ✓ *è integrata e consistente;*
- ✓ *è rappresentativa dell'evoluzione temporale;*
- ✓ *non volatile.*



Le interrogazioni

■ OLTP:

- ✓ Le interrogazioni eseguono transazioni che leggono e scrivono un ridotto numero di record da diverse tabelle legate da semplici relazioni.
- ✓ Il nucleo sostanziale del carico di lavoro è “congelato” all'interno dei programmi applicativi.

■ OLAP:

- ✓ Le interrogazioni effettuano un'analisi dinamica e multidimensionale che richiede la scansione di un'enorme quantità di record per calcolare un insieme di dati numerici di sintesi che quantifichino le prestazioni dell'azienda.
- ✓ L'interattività è una caratteristica irrinunciabile delle sessioni di analisi e fa sì che il carico di lavoro effettivo vari continuamente nel tempo.

Riassumendo:

	Database operazionali	Data warehouse
utenti	migliaia	centinaia
carico di lavoro	transazioni predefinite	interrogazioni di analisi <i>ad hoc</i>
accesso	a centinaia di record, in lettura e scrittura	a milioni di record, per lo pi in lettura
scopo	dipende dall'applicazione	supporto alle decisioni
dati	elementari, sia numerici sia alfanumerici	di sintesi, prevalentemente numerici
integrazione dei dati	per applicazione	per soggetto
qualit^	in termini di integrit^	in termini di consistenza
copertura temporale	solo dati correnti	dati correnti e storici
aggiornamenti	continui	periodici
modello	normalizzato	denormalizzato, multidimensionale
ottimizzazione	per accessi OLTP su una frazione del database	per accessi OLAP su gran parte del database
sviluppo	a cascata	iterativo



Verso il modello multidimensionale

“Che incassi sono stati registrati l’anno passato per ciascuna regione e ciascuna categoria di prodotto?”

“Che correlazione esiste tra l’andamento dei titoli azionari dei produttori di PC e i profitti trimestrali lungo gli ultimi 5 anni?”

“Quali sono gli ordini che massimizzano gli incassi?”

“Quale di due nuove terapie risulterà in una diminuzione della durata media di un ricovero?”

“Che rapporto c’è tra i profitti realizzati con spedizioni di meno di 10 elementi e quelli realizzati con spedizioni di più di 10 elementi?”



Il modello multidimensionale

- È il fondamento per la rappresentazione e l'interrogazione dei dati nei data warehouse.
- I *fatti* di interesse sono rappresentati in *cubi* in cui:
 - ✓ ogni cella contiene *misure* numeriche che quantificano il fatto da diversi punti di vista;
 - ✓ ogni asse rappresenta una *dimensione* di interesse per l'analisi;
 - ✓ ogni dimensione può essere la radice di una *gerarchia* di attributi usati per aggregare i dati memorizzati nei cubi base.

Il cubo delle vendite

