

BUSINESS INTELLIGENCE E DATA WAREHOUSING: INTRODUZIONE

Vedremo



- **Cosa si intende per Business Intelligence**
- **Come i dati possono fornire supporto alle decisioni**
- **Gli elementi di un sistema di Business Intelligence**
- **Il Data Warehouse**
- **Preparazione dei dati per l'analisi e qualità dei dati**

Preliminari: dati e informazioni

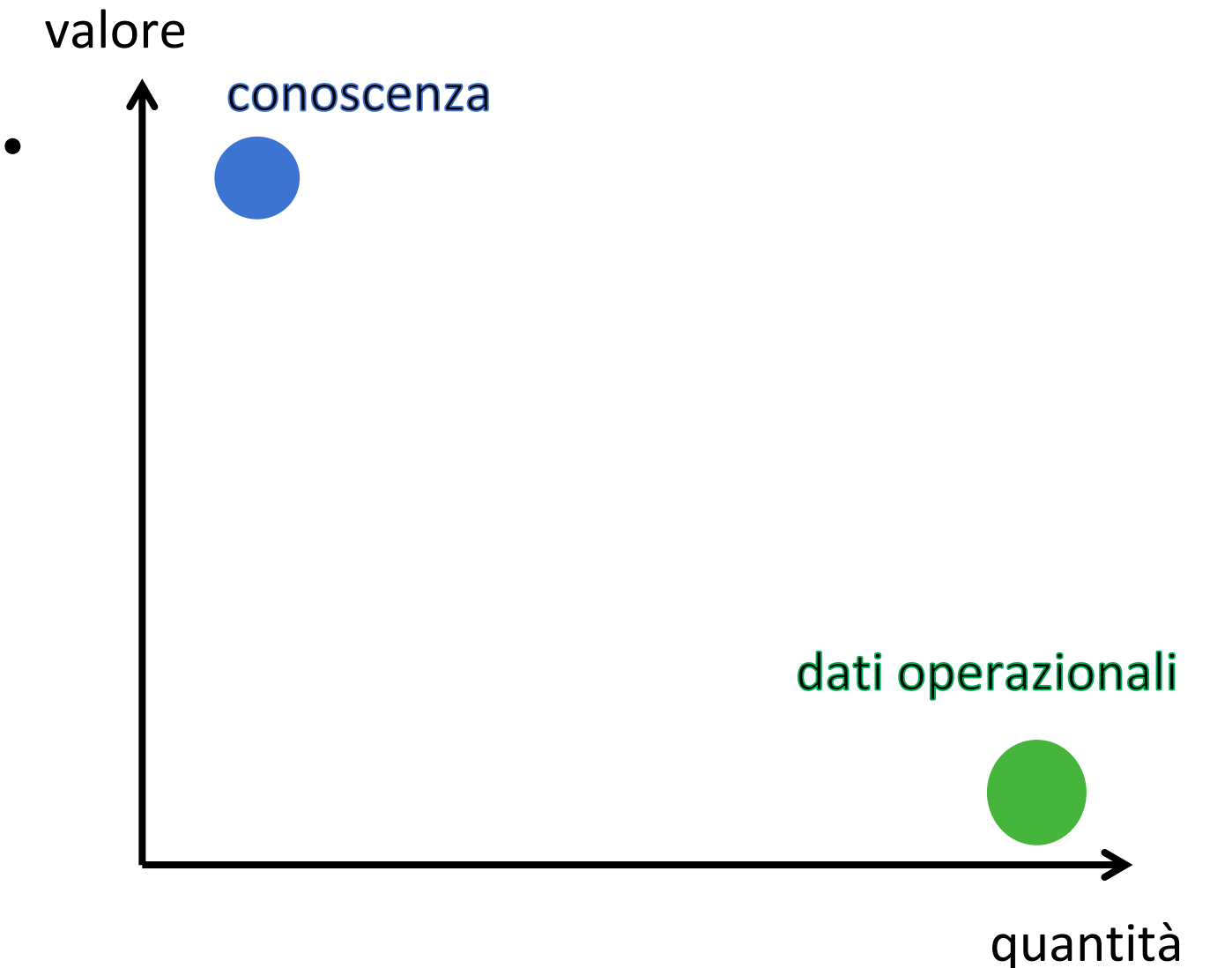


- **Dati:** codifica strutturata di singole entità primarie o di transazioni che le riguardano:
 - ad esempio, le descrizioni e i prezzi dei prodotti, gli importi sugli scontrini di vendita.
- **Informazioni:** dotate di significato per chi le riceve in uno specifico contesto;
 - dato + contesto interpretativo.
- **Informazioni aggregate:** ottenute mediante elaborazione e sintesi a partire da dati elementari.
- **Conoscenza:** informazione utilizzata per prendere decisioni.

Valore strategico delle informazioni



- Tutte le organizzazioni hanno a disposizione una grande mole di dati operazionali, cioè raccolti e gestiti per garantire la loro operatività quotidiana.
- I dati operazionali sono disponibili in grandissima quantità, ma hanno un basso valore informativo e strategico.
- La Business Intelligence si basa sul distillare la **conoscenza**, cioè le informazioni strategiche, o utili a fini decisionali, presenti nei dati operazionali.



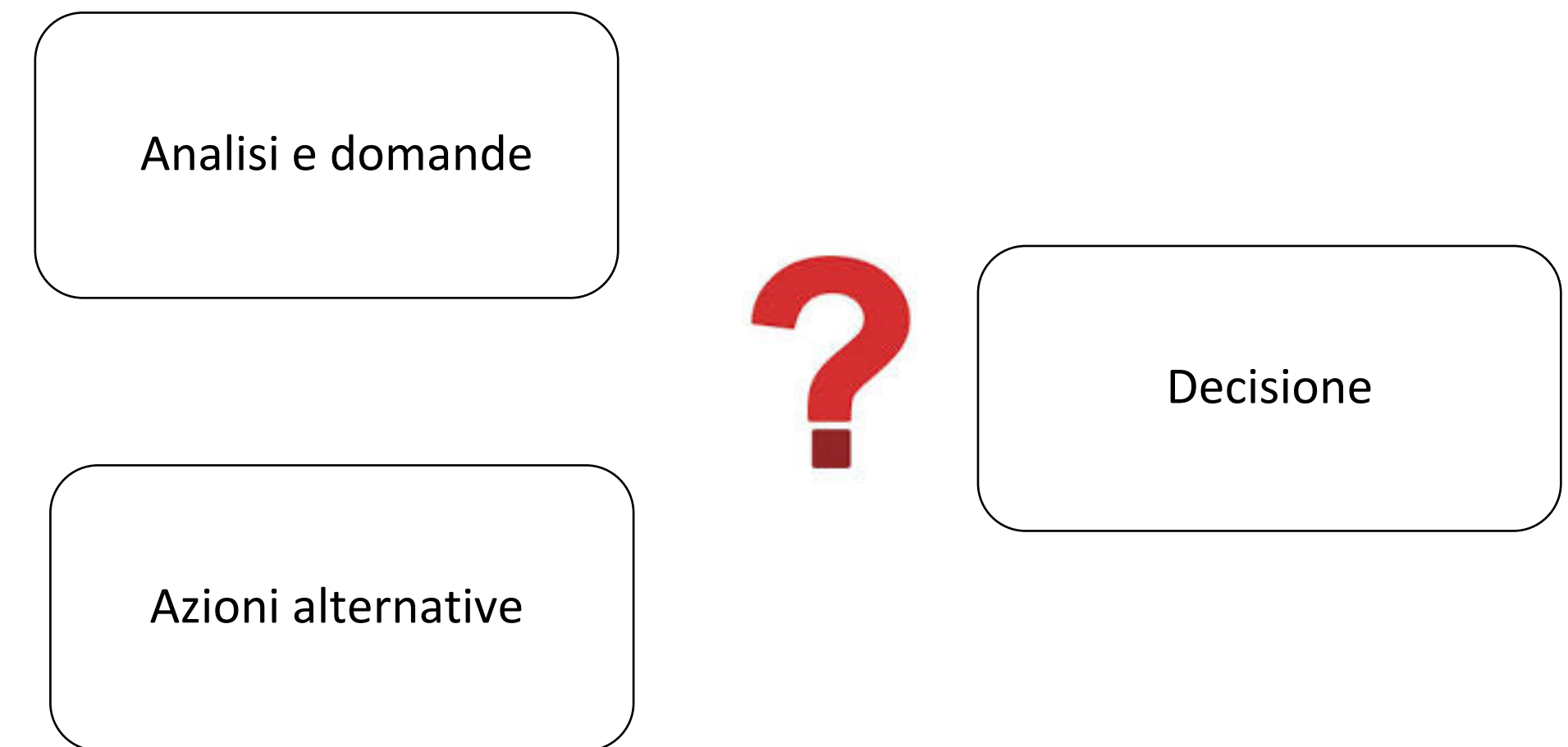
Business Intelligence



Business Intelligence = un insieme di strumenti e tecniche che danno la possibilità a un'organizzazione di trasformare i propri dati in conoscenza che consenta di prendere decisioni efficaci e tempestive.

Tali strumenti permettono di:

- considerare molte alternative;
- giungere a conclusioni più accurate.



Business Intelligence e decisioni basate sui dati



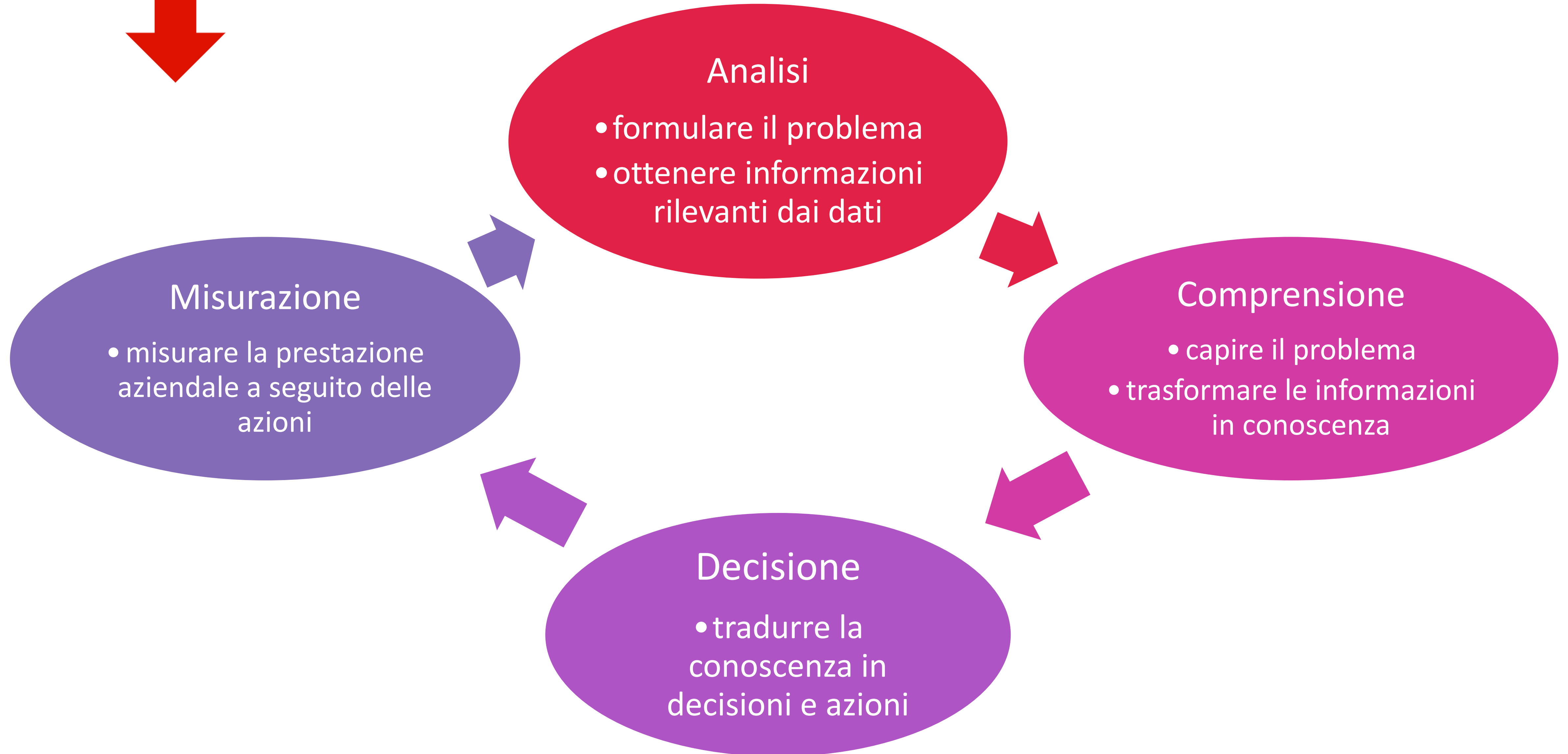
I sistemi di Business Intelligence sono utilizzati dai decisori:

- per ottenere una conoscenza completa del business e dei fattori che lo influenzano;
- per definire e supportare le proprie strategie di business.

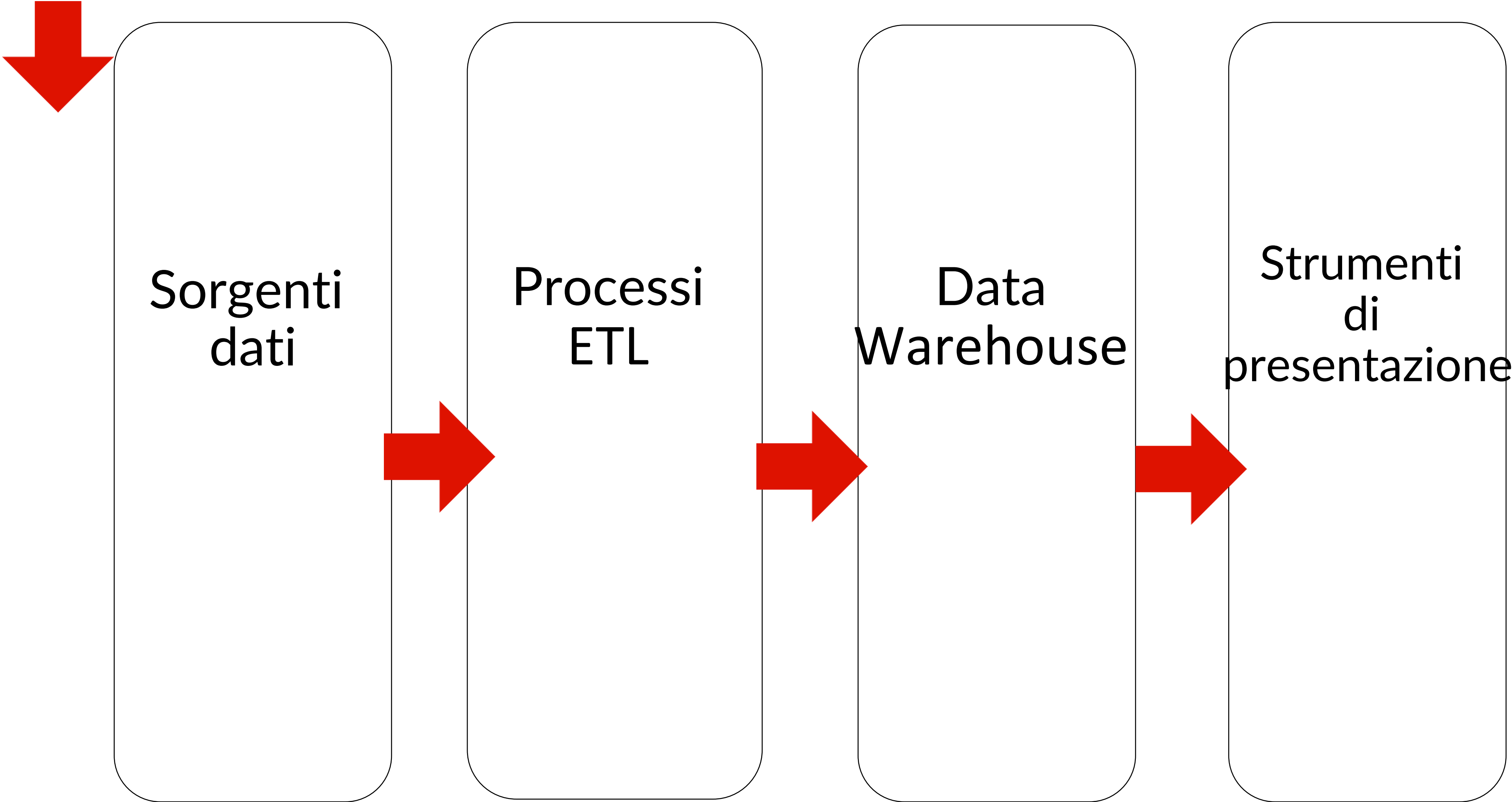
L'obiettivo è consentire decisioni basate sui dati.

La conoscenza estratta dai dati permette di controllare meglio il rischio insito in ogni decisione.

Ciclo decisionale



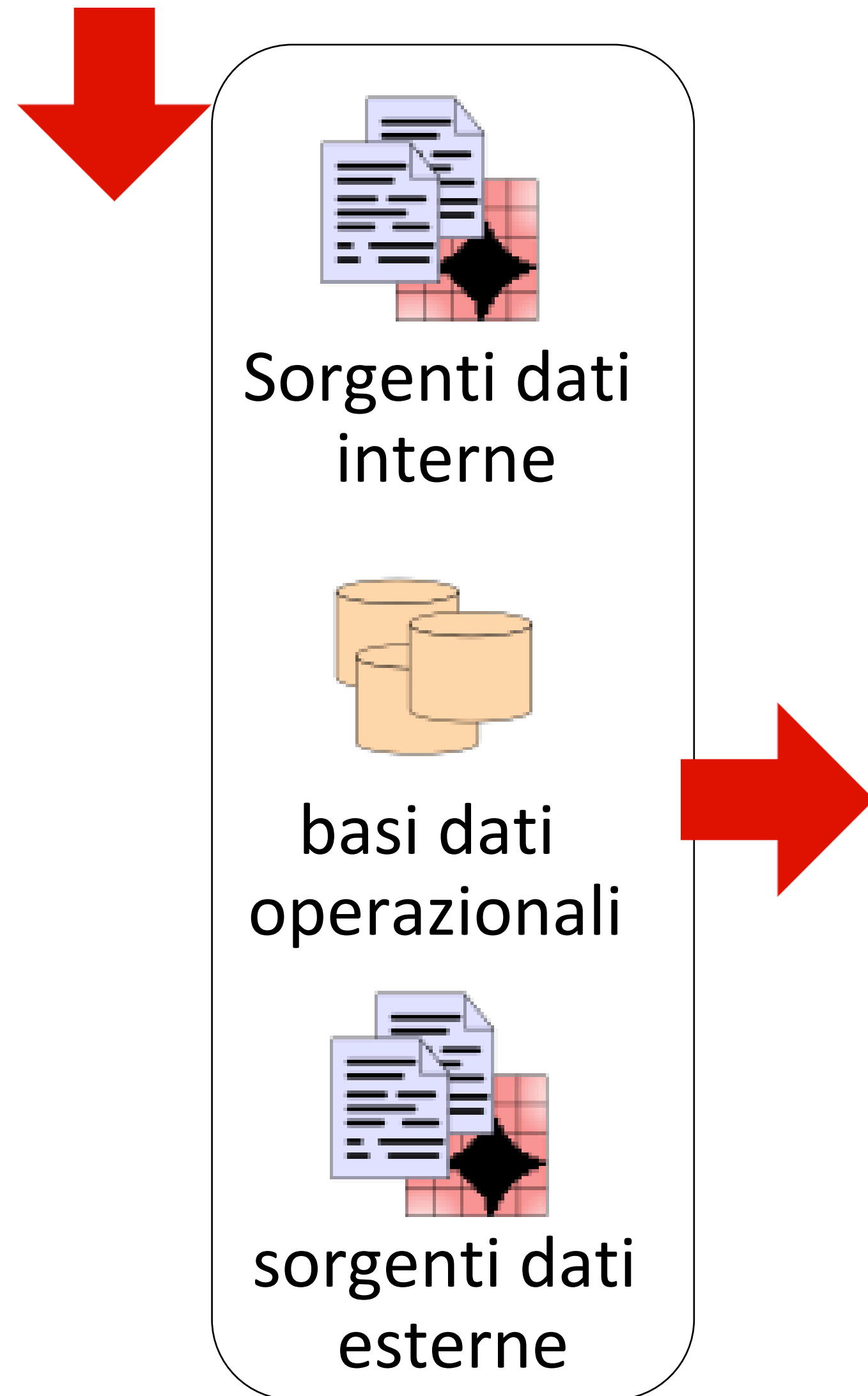
Componenti di un sistema di Business Intelligence



Back-end

Front-end

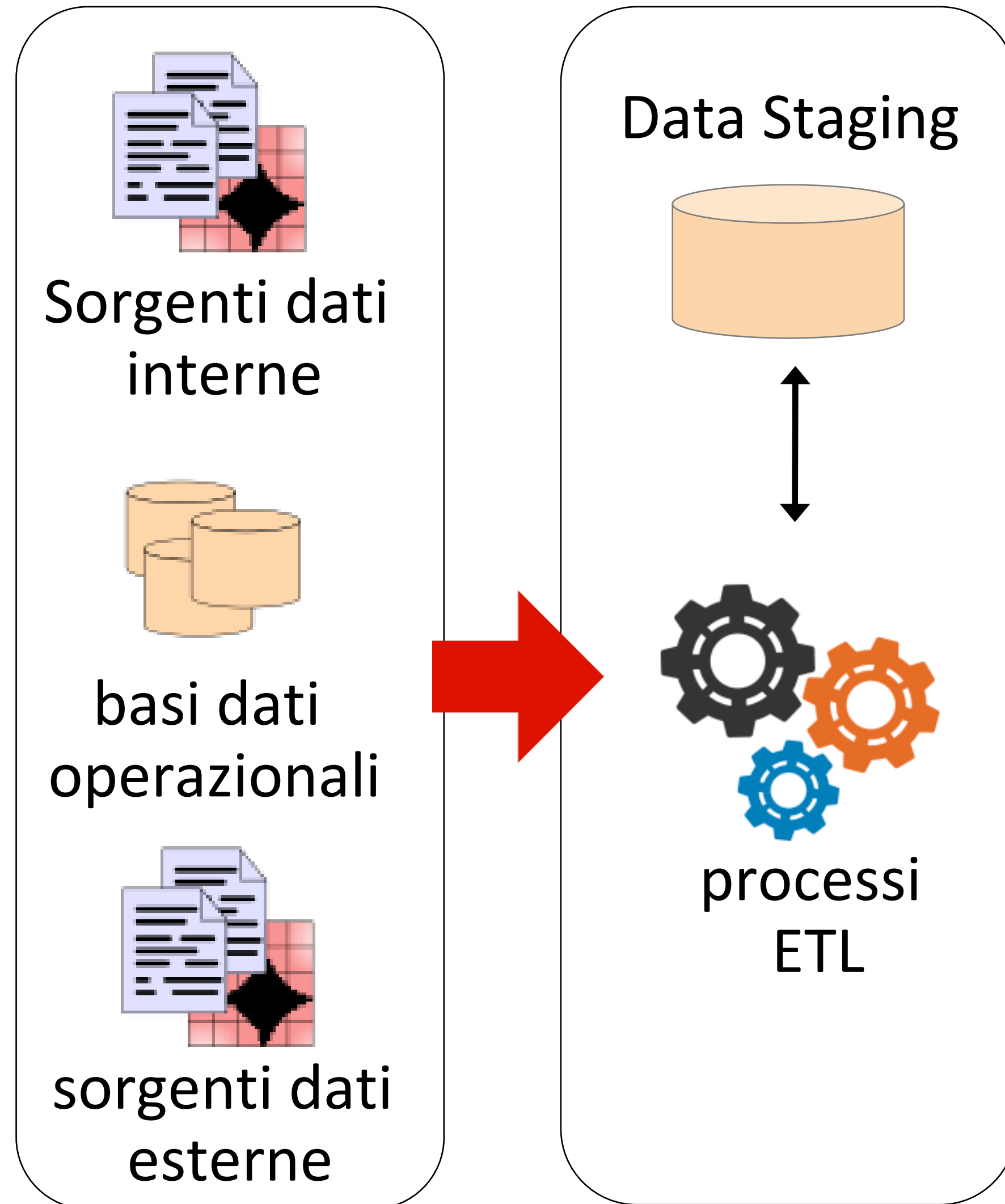
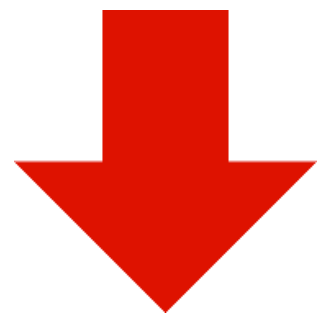
Sorgenti dati



Sorgenti dati

- Dati operazionali, o di dettaglio
 - interni o
 - esterni all'organizzazione.
- Diversi livelli di strutturazione:
 - Basi di dati operazionali;
 - Dati semi-strutturati;
 - Fogli di calcolo;
 - File di testo;
 - ...

Processi ETL



Sorgenti dati

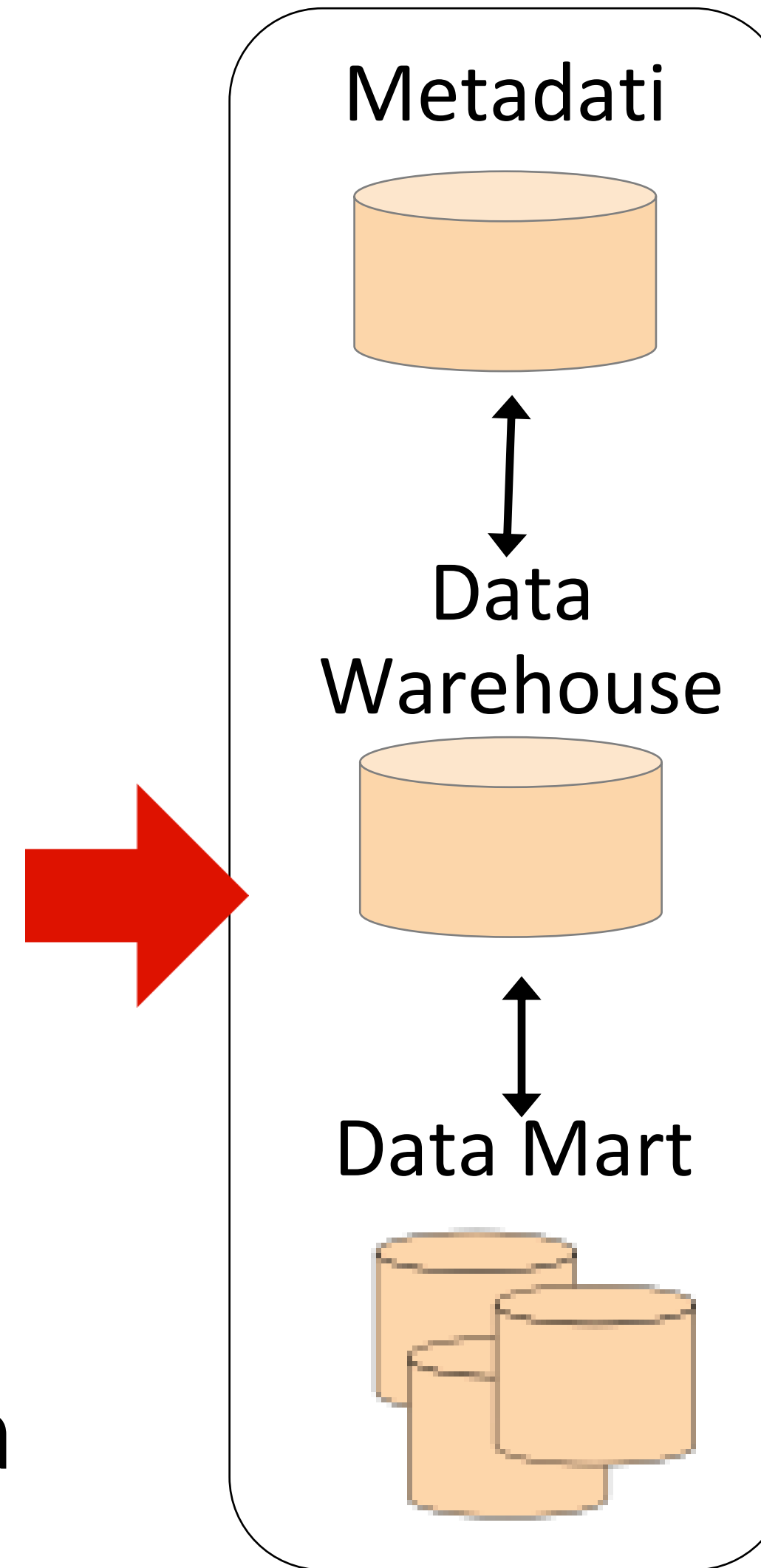
ETL

- I processi ETL alimentano il livello di analisi dei dati.
- **E**: estrazione dei dati dalle sorgenti operazionali;
- **T**: workflow che automatizza tutte le trasformazioni necessarie (integrazione e riconciliazione, pulizia, adattamento di formato);
- **L**: caricamento dei dati trasformati nel repository su cui verranno effettuate le analisi.

Data Warehouse

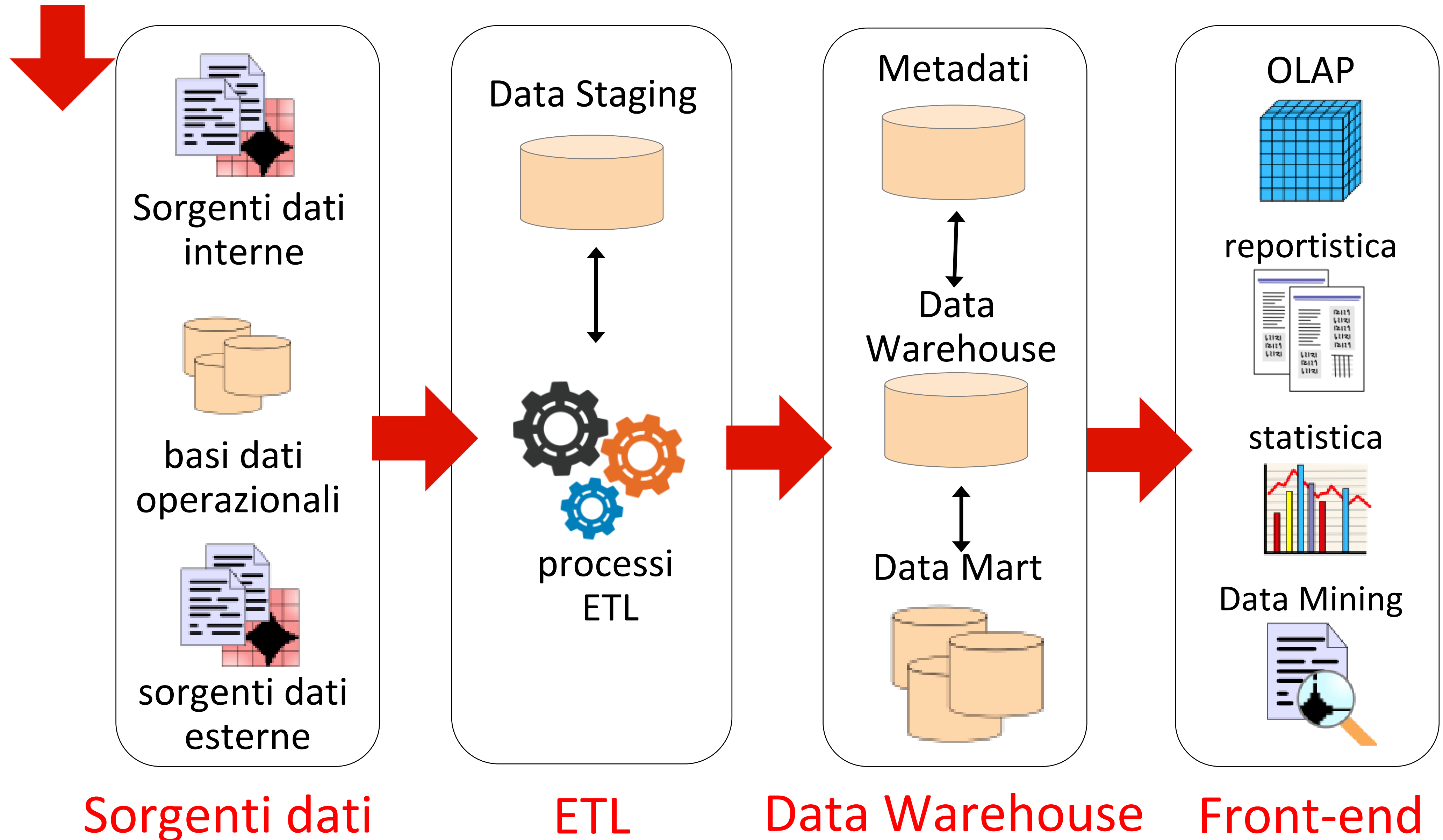


- Data Warehouse = repository di informazioni che raccoglie e integra dati provenienti da fonti diverse ed eterogenee rendendoli disponibili per analisi finalizzate alla pianificazione e al processo decisionale.
- Data Mart = sottoinsieme o aggregazione dei dati archiviati in un Data Warehouse primario, include informazioni rilevanti per una specifica area o reparto aziendale, categoria di utenti o linea di business.



Data Warehouse

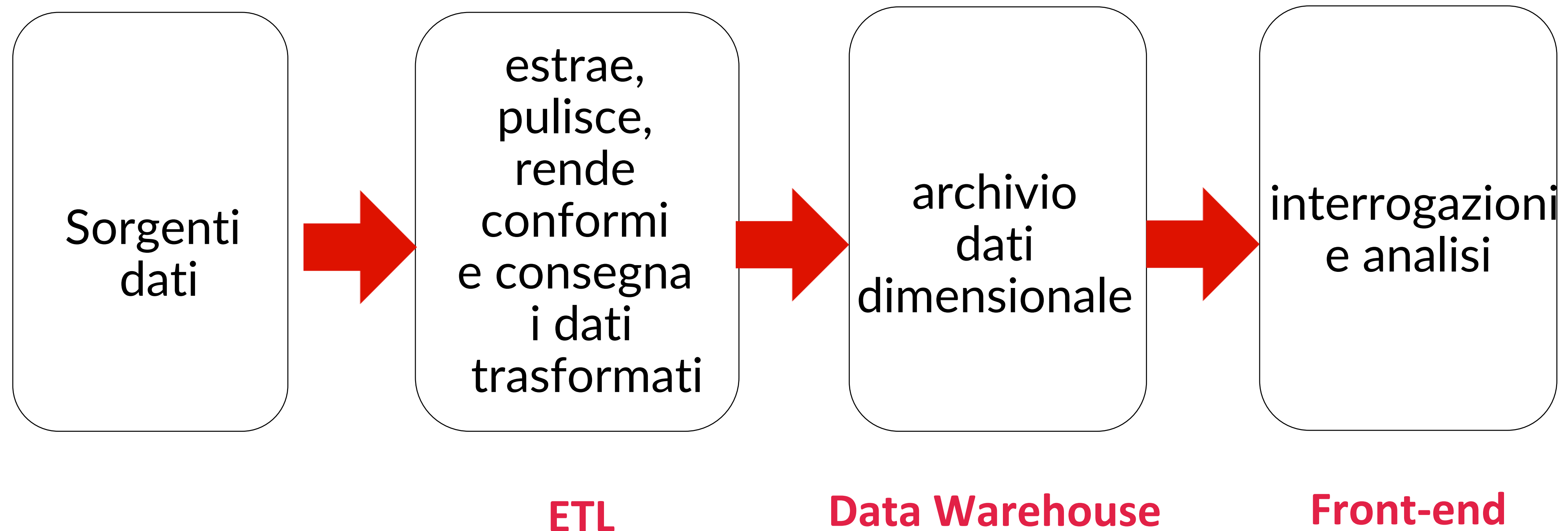
Strumenti di presentazione



Business Intelligence & Data Warehousing



- Sistema che estrae, pulisce, rende conformi i dati sorgente e li consegna a un archivio dati dimensionale, abilitando e implementando interrogazioni e analisi finalizzate a prendere decisioni.



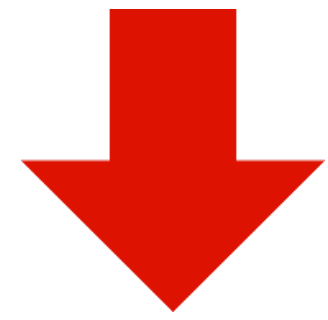
Data Warehouse - Caratteristiche



Un Data Warehouse è una raccolta di dati che supporta i processi decisionali:

- orientato al soggetto;
- integrato e consistente;
- mostra l'evoluzione dei dati nel tempo;
- non è volatile.

OLAP vs OLTP



Perché c'è bisogno di un altro repository?

- Le esigenze operazionali (OLTP) sono molto diverse da quelle analitiche (OLAP);
- gestirle entrambe in un unico repository non soddisfa adeguatamente nessuna delle due.

OLTP

Online Transactional Processing

Esempi:

- transazione di vendita;
- disponibilità di un prodotto in un punto vendita;
- prezzo di un prodotto.

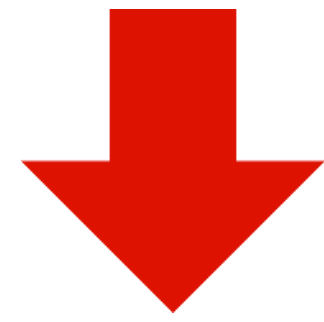
OLAP

Online Analytical Processing

Esempi:

- confronto incassi mensili di punti vendita;
- andamento stagionale delle vendite di latticini.

Differenze tra un Data Warehouse e un database



	Basi di dati operazionali	Data Warehouse
scopo	application-dependent	supporto alle decisioni
focus	sulle applicazioni	sui soggetti
qualità	in termini di integrità	in termini di consistenza
dati	elementari, numerici e alfanumerici	aggregati, principalmente numerici
workload	transazioni predefinite	interrogazioni di analisi ad hoc
utenti	migliaia	centinaia

Differenze tra un Data Warehouse e un database (2)



	Basi di dati operazionali	Data Warehouse
accesso	in lettura e scrittura, a centinaia di record	principalmente in sola lettura, a milioni di record
ottimizzato per	accessi OLTP su una frazione limitata dei dati	accessi OLAP che coinvolgono una grande frazione dei dati
aggiornamento	continuo	periodico
copertura temporale	dati attuali	dati storici

Differenze tra OLAP e OLTP - Conseguenze



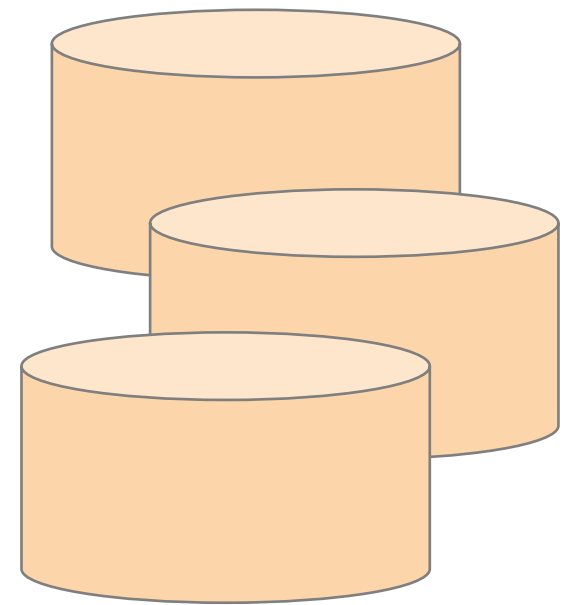
- Alcuni dei cardini dell'organizzazione e della gestione dei dati a fini operazionali vanno quindi rivisti:
- meccanismi transazionali (non più necessari);
- normalizzazione degli schemi;
- indici (per accessi selettivi).

Anche il modello stesso di organizzazione dei dati può essere esplicitamente finalizzato all'analisi: **modello dimensionale**.

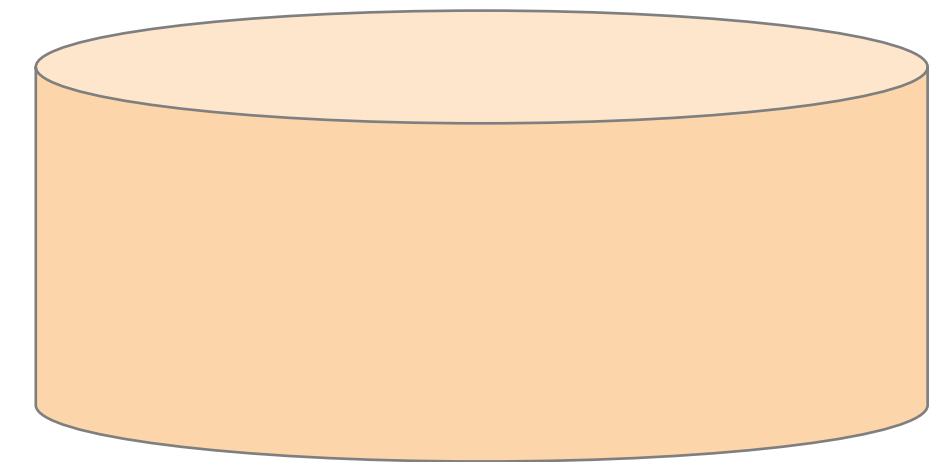
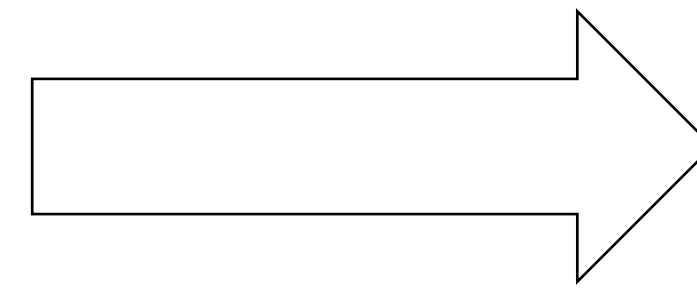
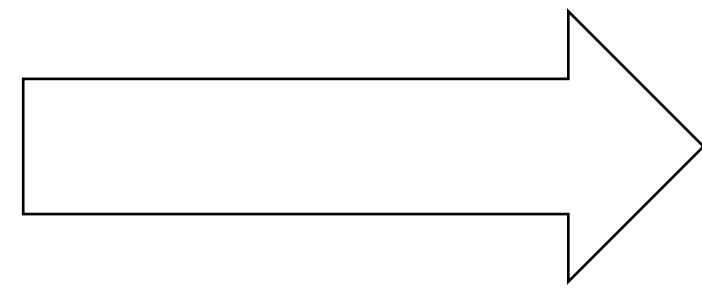
Processi ETL



- trasformano i dati dal formato sorgente (operazionale) a quello di destinazione (analitico);
- quando il repository di destinazione viene popolato per la prima volta e ogni volta che viene aggiornato regolarmente.



Sorgenti (operazionali)



Destinazione (analitica)

Perché trasformare?



- Per poter estrarre conoscenza utile i dati devono essere integrati e di buona qualità.
- Avere fonti dati operative diverse ed eterogenee richiede di stabilire una mappatura tra il livello di dati di origine e il livello di analisi di destinazione affrontando due problematiche:
 - 1) Eterogeneità (pulizia e riconciliazione);
 - 2) Passaggio da dettaglio a aggregato, dal modello dei dati operativo a quello analitico (multidimensionale).

Qualità dei dati



La prima «metà» della trasformazione include anche una fase di pulizia per migliorare la qualità dei dati. I problemi di qualità possono essere presenti anche nelle singole sorgenti, ma sono amplificati dall'integrazione di sorgenti eterogenee:

- dati mancanti;
- dati duplicati;
- dati impossibili o errati;
- uso imprevisto dei campi;
- valori incoerenti per una singola entità perché sono state utilizzate diverse rappresentazioni o pratiche diverse;
- valori incoerenti per una singola entità a causa di errori di battitura;
- valori incoerenti che sono logicamente associati.

Come trasformare?



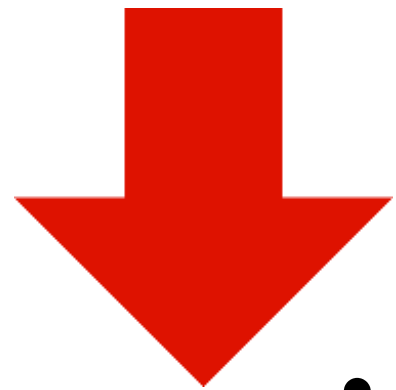
Molte alternative:

- Linguaggi di scripting, SQL procedurale, ...
- Strumenti di ETL;
- ELT vs ETL (la trasformazione avviene all'interno del sistema che gestisce il repository analitico).

Gli strumenti ETL forniscono supporto per:

- Definizione di una pipeline (connessione alla sorgente dati, workflow di trasformazioni, alimentazione della destinazione);
- Scheduling e monitoraggio dell'esecuzione automatizzata della pipeline.

Riepilogo e conclusioni finali



Abbiamo visto:

- **Business Intelligence usa i dati, trasformandoli in conoscenza, per aiutare a prendere decisioni**
- **Sistemi di Business Intelligence, basati su componenti di back-end e front-end**
- **Il Data Warehouse come repository integrato per l'analisi**
- **Processi ETL per popolare il Data Warehouse a partire dalle sorgenti operazionali e migliorare la qualità dei dati.**