

Architetture Evolute nei Sistemi Informativi

Scalabilità delle Applicazioni

- **carico:**
insieme di tutte le applicazioni (query)
- **scalabilità:**
abilità di conservare prestazioni elevate al crescere del carico
- **dimensioni di crescita:**
 - numero delle query
 - complessità delle query

Due tipologie di carico

- **transazionale**

carico: transazioni di update

misura: tps (transazioni per secondo)

tempo di risposta: pochi secondi

- **analisi dei dati**

carico: query SQL complessa

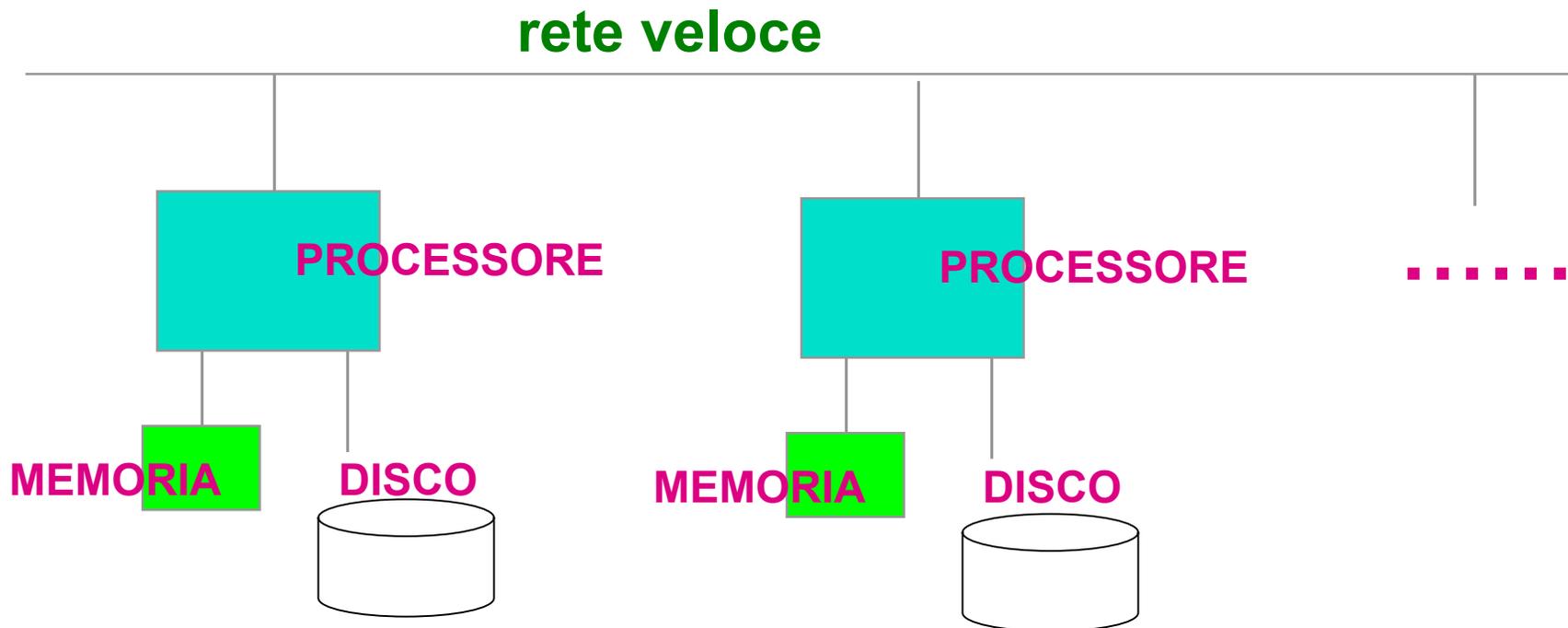
tempo di risposta: variabile

Parallelismo

- ottenuto tramite molti processori che cooperano in una unica architettura informatica
- esistono due tipi di **parallelismo**
 - inter-query**
ciascuna query affidata ad un solo processore (per carichi transazionali)
 - intra-query**
ciascuna query affidata a molti processori (per carichi di analisi dei dati)

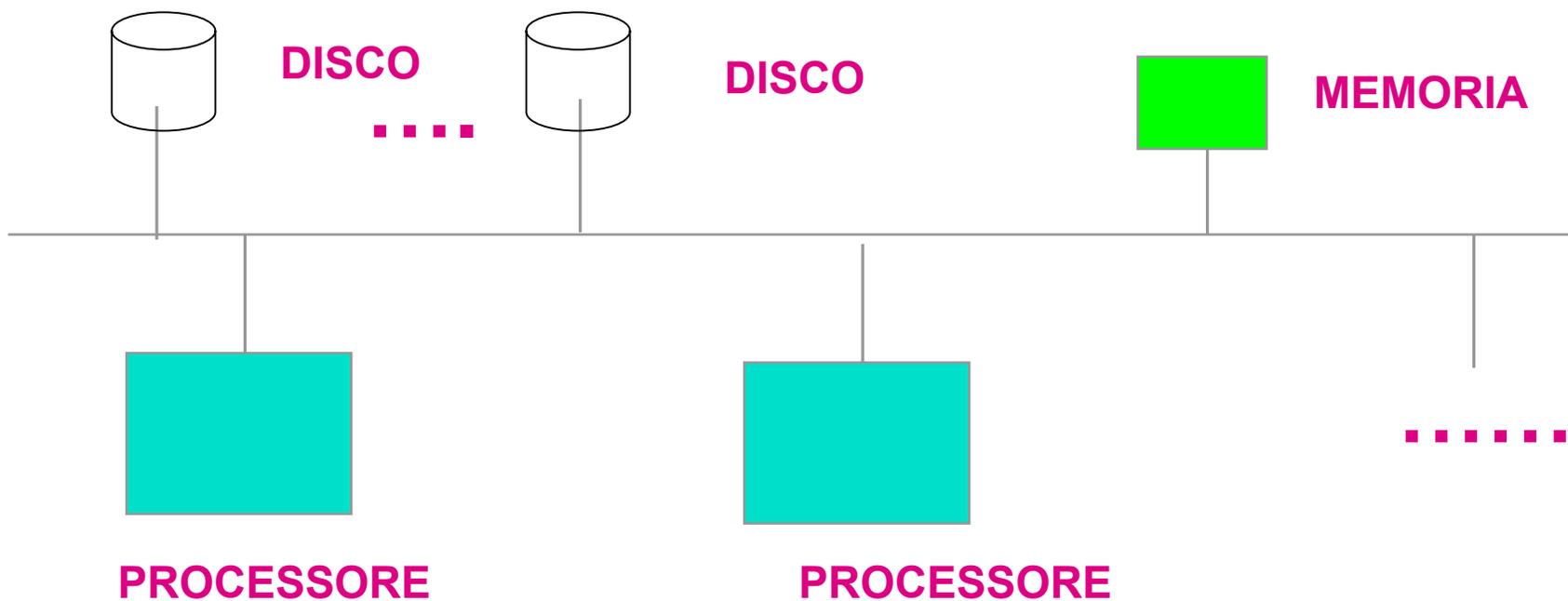
Architetture parallele a confronto

SHARED-NOTHING



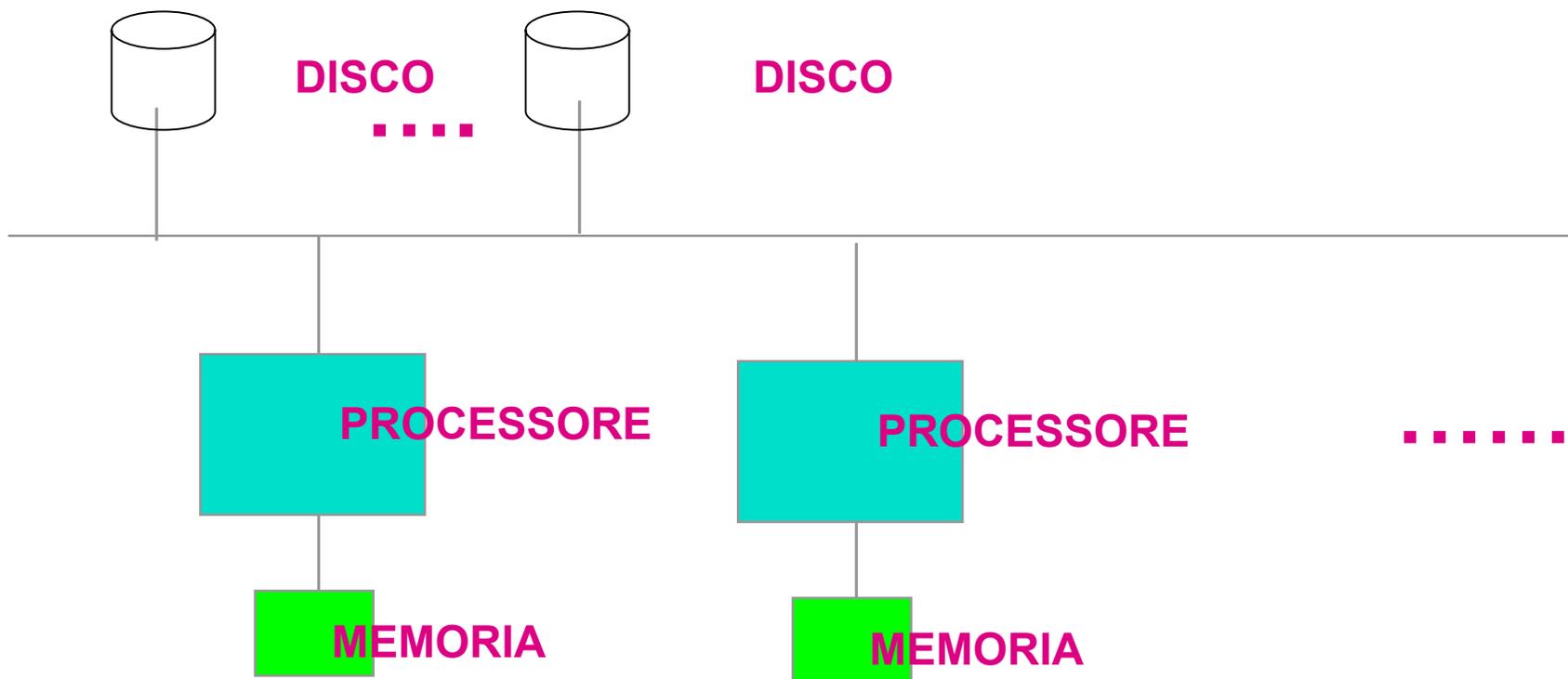
Architetture parallele a confronto

SHARED-MEMORY



Architetture parallele a confronto

SHARED-DISKS



La scelta vincente per un DBMS

SHARED-NOTHING

- è l'architettura più flessibile
- non presenta “colli di bottiglia”
 - memoria e bus per comunicazione e coordinamento processori in SHARED-MEMORY
 - dischi per accesso ai dati (!) in SHARED-DISK

Benchmark

metodi per confrontare le prestazioni di sistemi diversi (in competizione)

standardizzazione:

1 del database

2 del carico

- codice transazioni
- modalità di invio
- frequenza di arrivo

3 della modalità di misurazione

• varie tipologie di carico

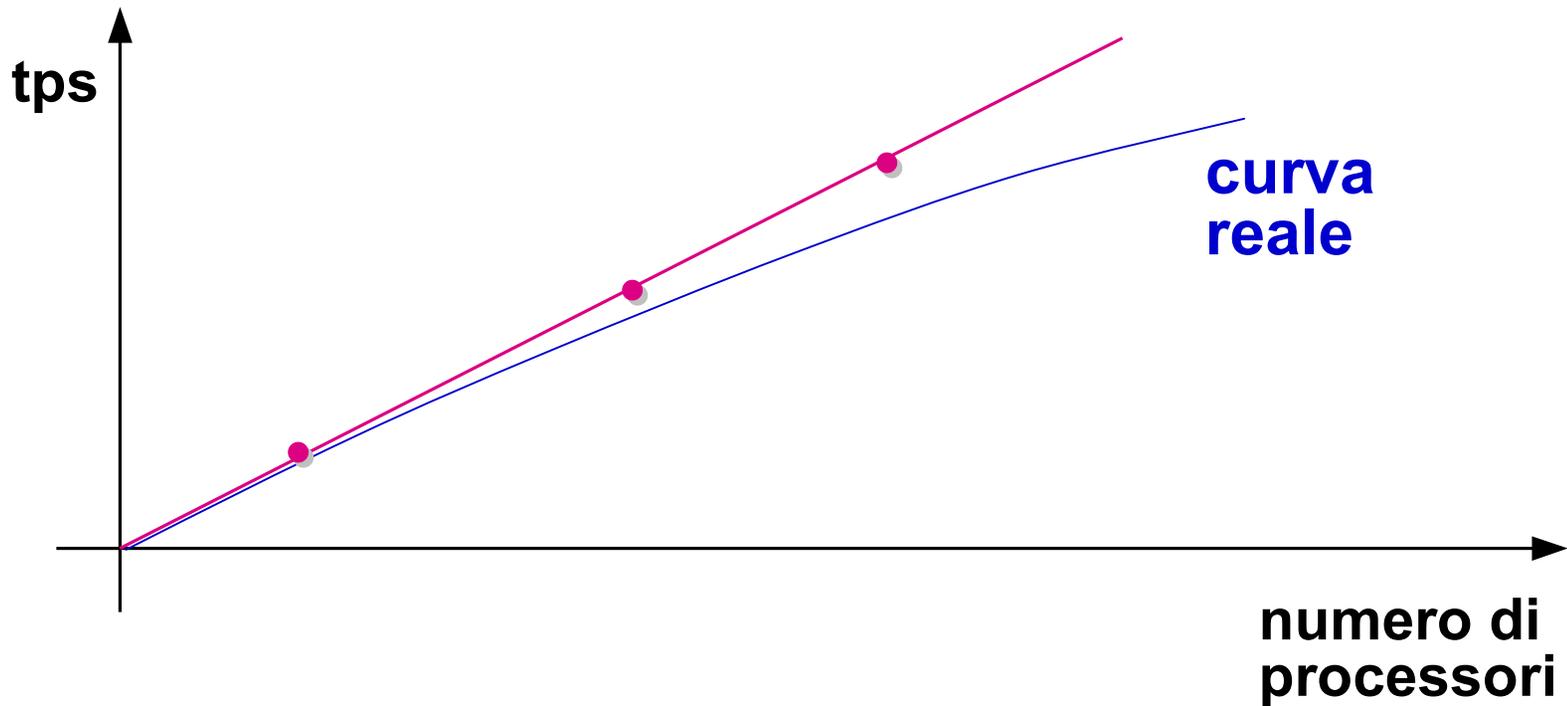
tpc-a: transazionale

tpc-b: misto

tpc-c: analisi dei dati

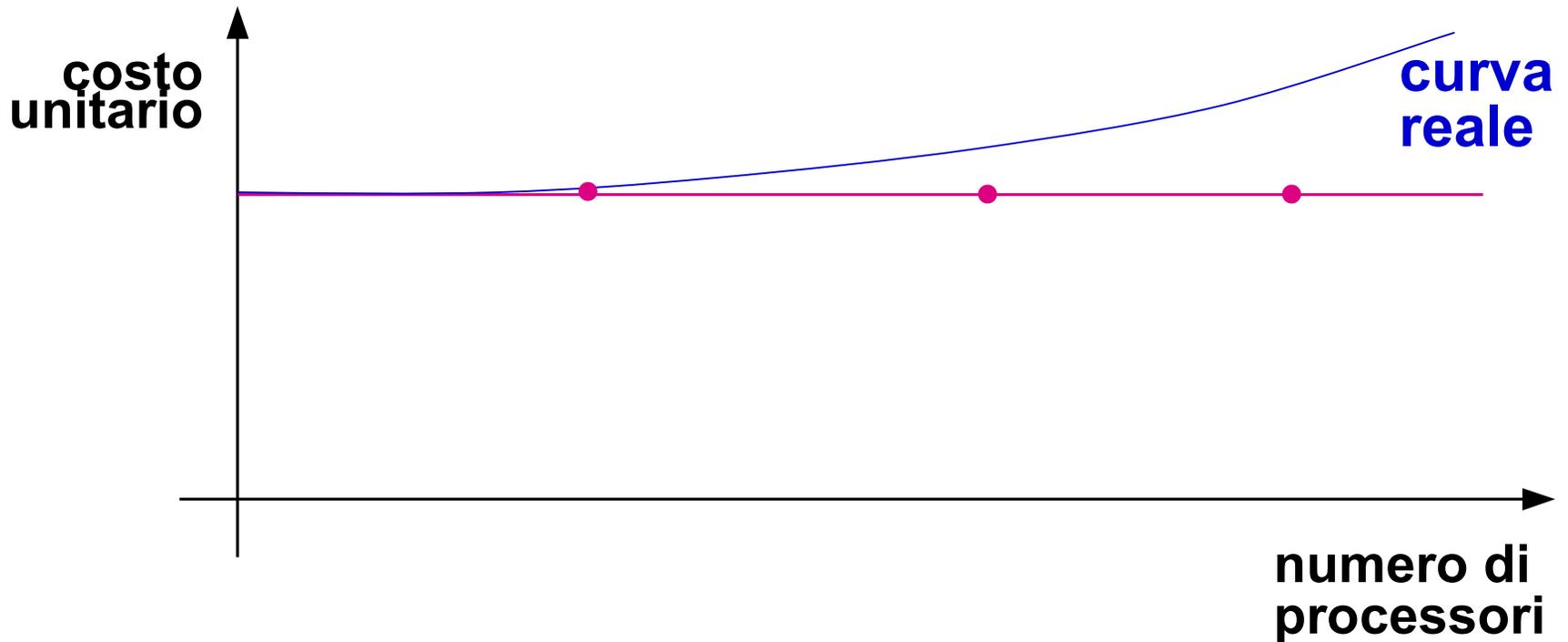
Curva di speed-up

misura il crescere di efficienza al crescere del numero di processori



Curva di scale-up

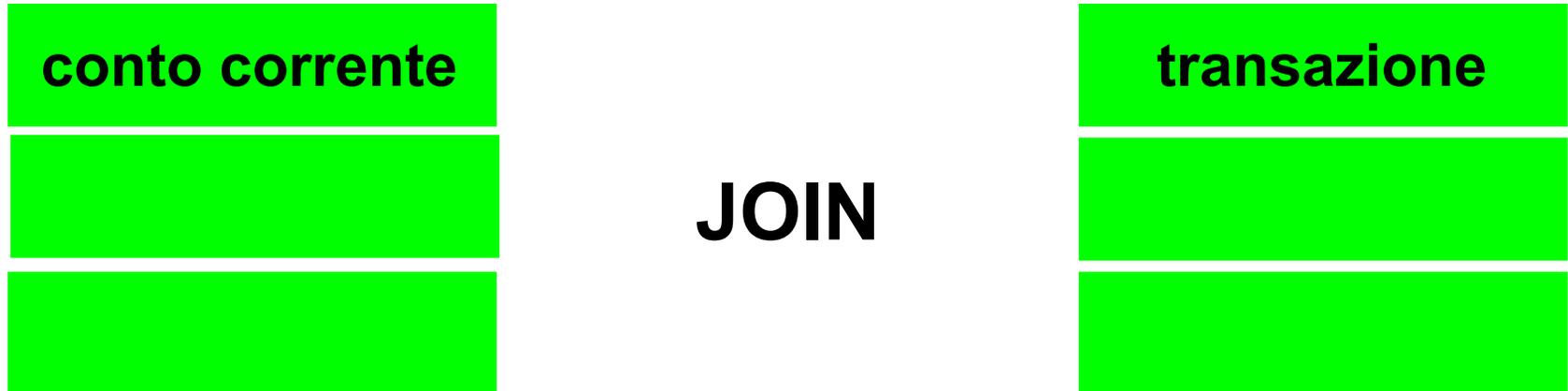
misura il crescere di costo unitario complessivo per transazione al crescere del numero di processori



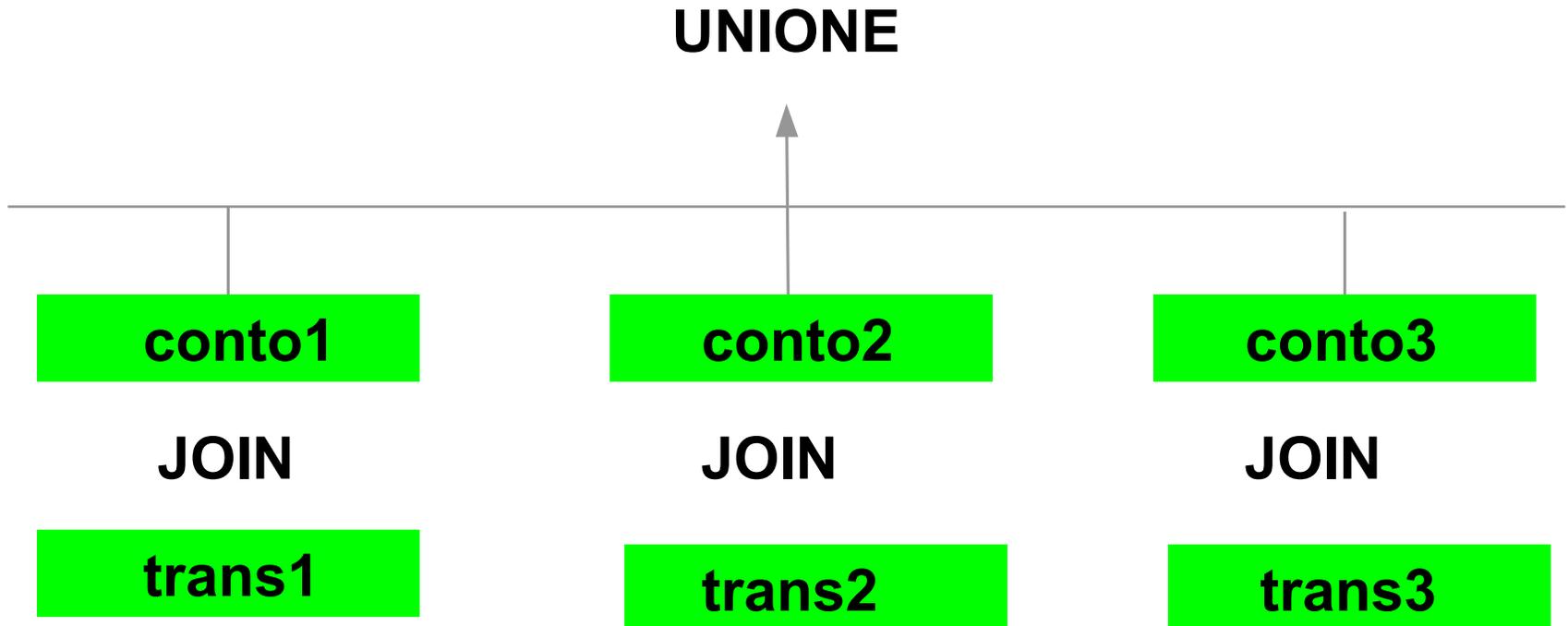
Join distribuito

è l'operazione di analisi dei dati
più onerosa

consideriamo:



Join distribuito



Replicazione dei dati

è un ingrediente fondamentale dei sistemi informativi

motivazioni:

- **efficienza**
- **affidabilità**
- **autonomia**
- **controllo**

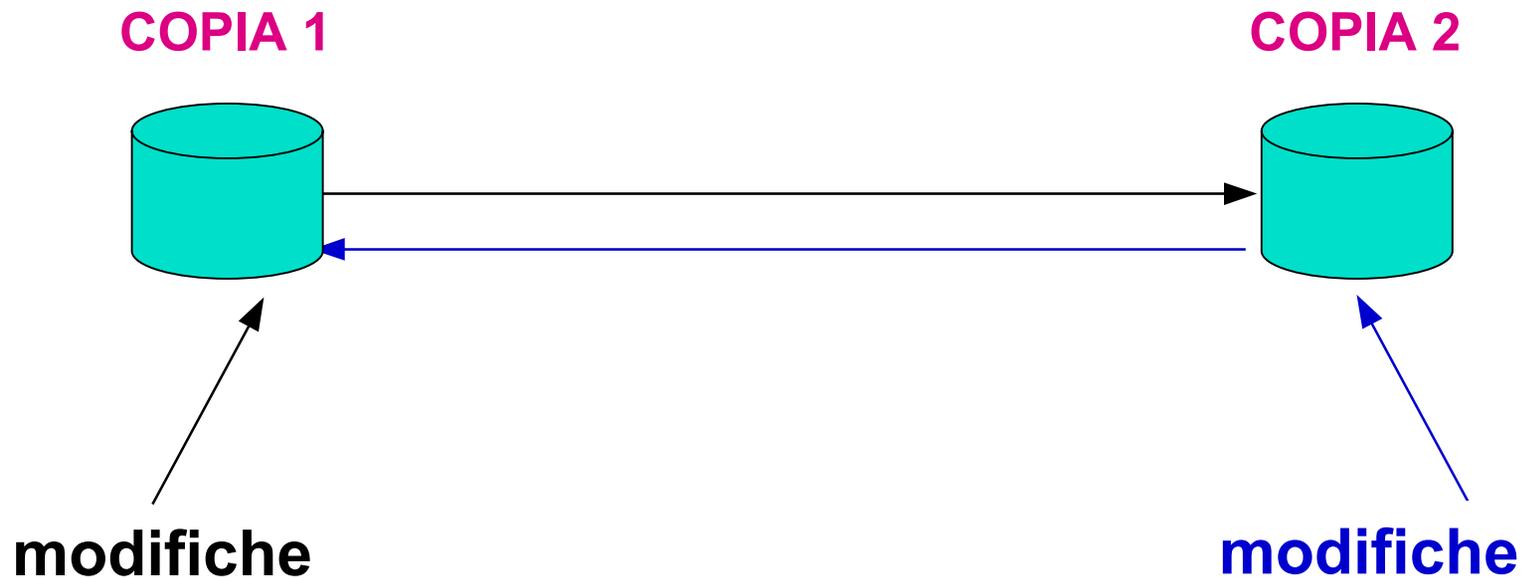
Modalità di replicazione

- **asimmetrica**



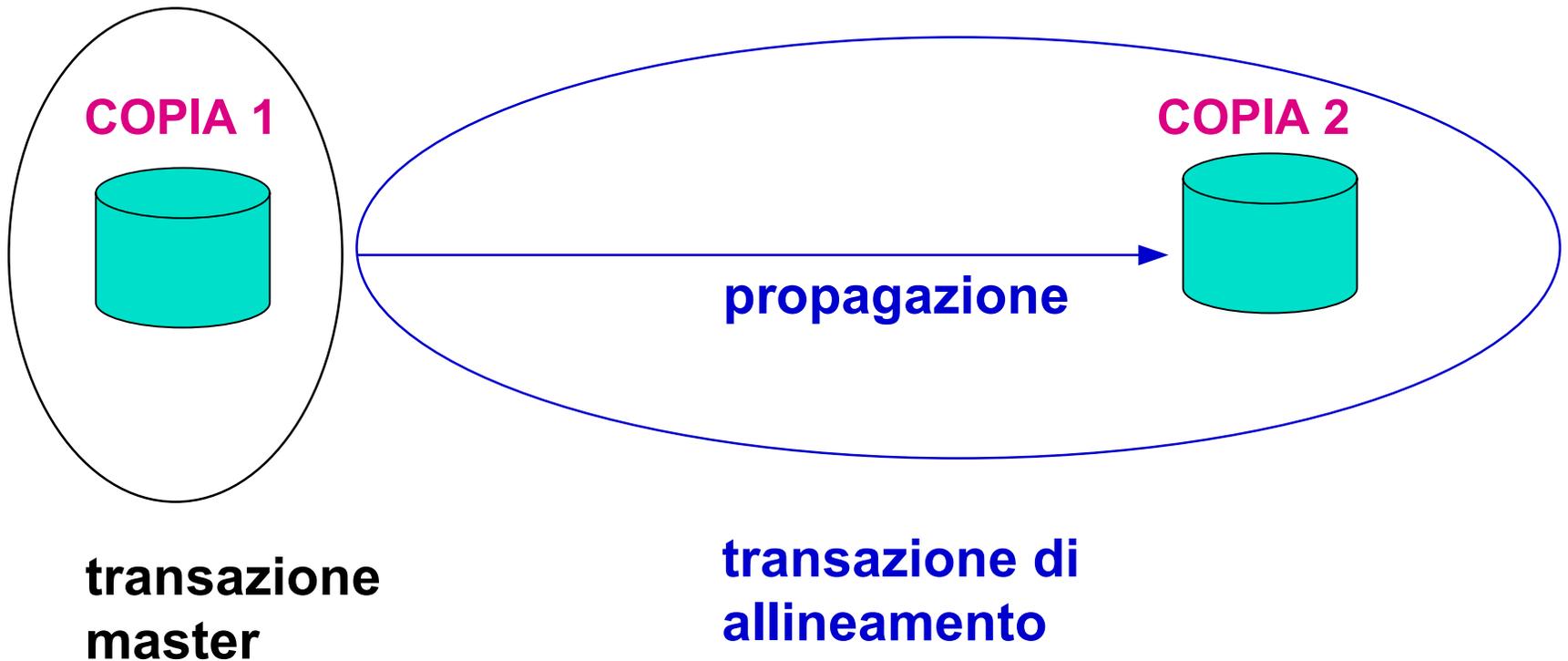
Modalità di replicazione

- **simmetrica**



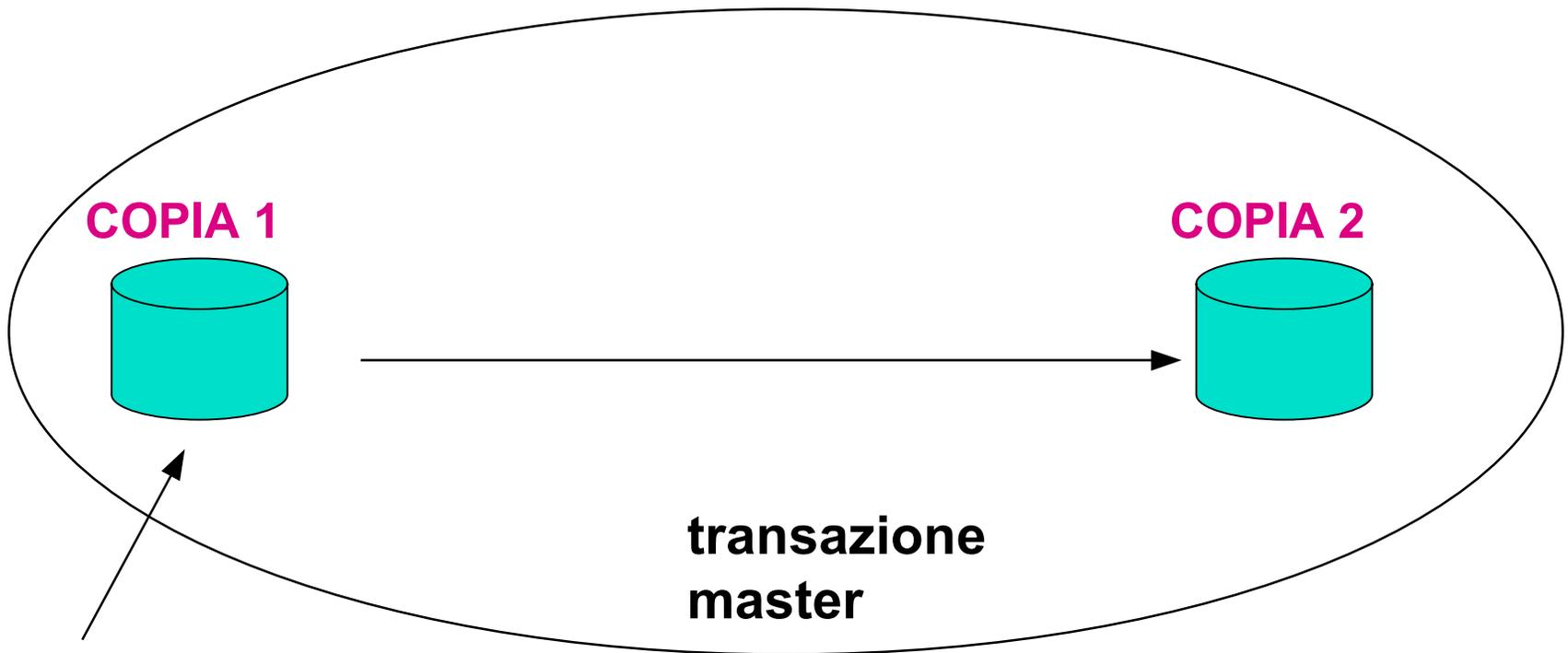
Modalità di trasmissione delle variazioni

- **trasmissione asincrona**



Modalità di trasmissione delle variazioni

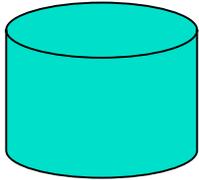
- **trasmissione sincrona**



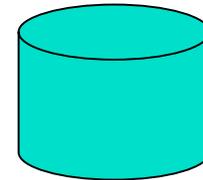
Modalità di allineamento

- refresh
- incrementale

COPIA 1



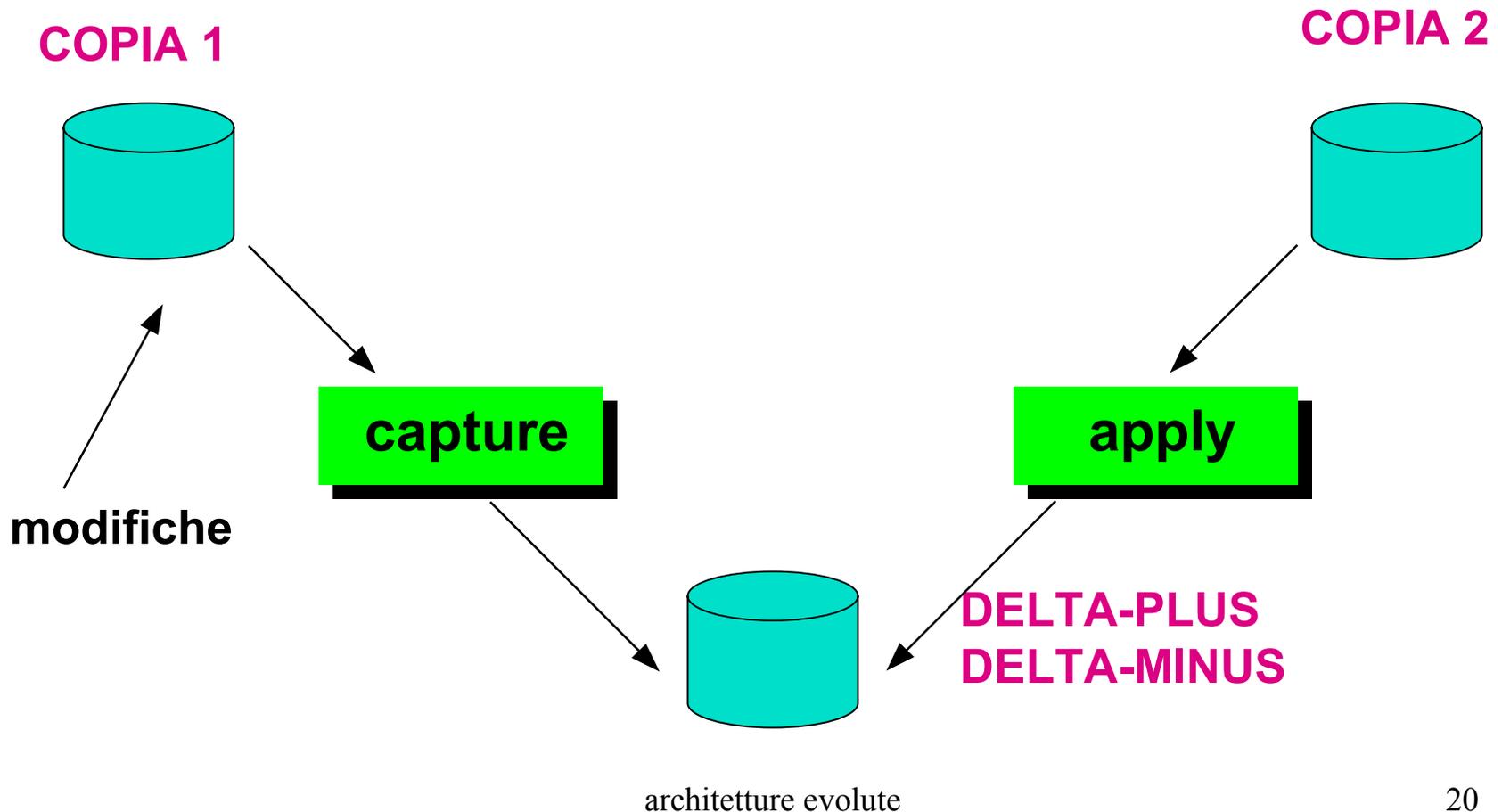
COPIA 2



DELTA-PLUS
DELTA-MINUS

Meccanismi per la replica asimmetrica, asincrona e incrementale

due moduli : **CAPTURE**, **APPLY**



Trigger di replicazione

**catturano le variazioni ai dati nelle
tabelle **DELTA-PLUS** e **DELTA-MINUS**
in modo trasparente alle applicazioni**

Trigger di replicazione

```
CREATE TRIGGER CAPTURE-INS  
AFTER INSERT ON PRIMARY  
FOR EACH ROW  
INSERT INTO DELTA-PLUS VALUES (NEW.*)
```

```
CREATE TRIGGER CAPTURE-DEL  
AFTER DELETE ON PRIMARY  
FOR EACH ROW  
INSERT INTO DELTA-MINUS VALUES (OLD.*)
```

```
CREATE TRIGGER CAPTURE-UPD  
AFTER UPDATE ON PRIMARY  
FOR EACH ROW  
BEGIN  
INSERT INTO DELTA-PLUS VALUES (NEW.*)  
INSERT INTO DELTA-MINUS VALUES (OLD.*)  
END
```

Applicazione

a comando

- **periodica**
- **guidata dalle modifiche**

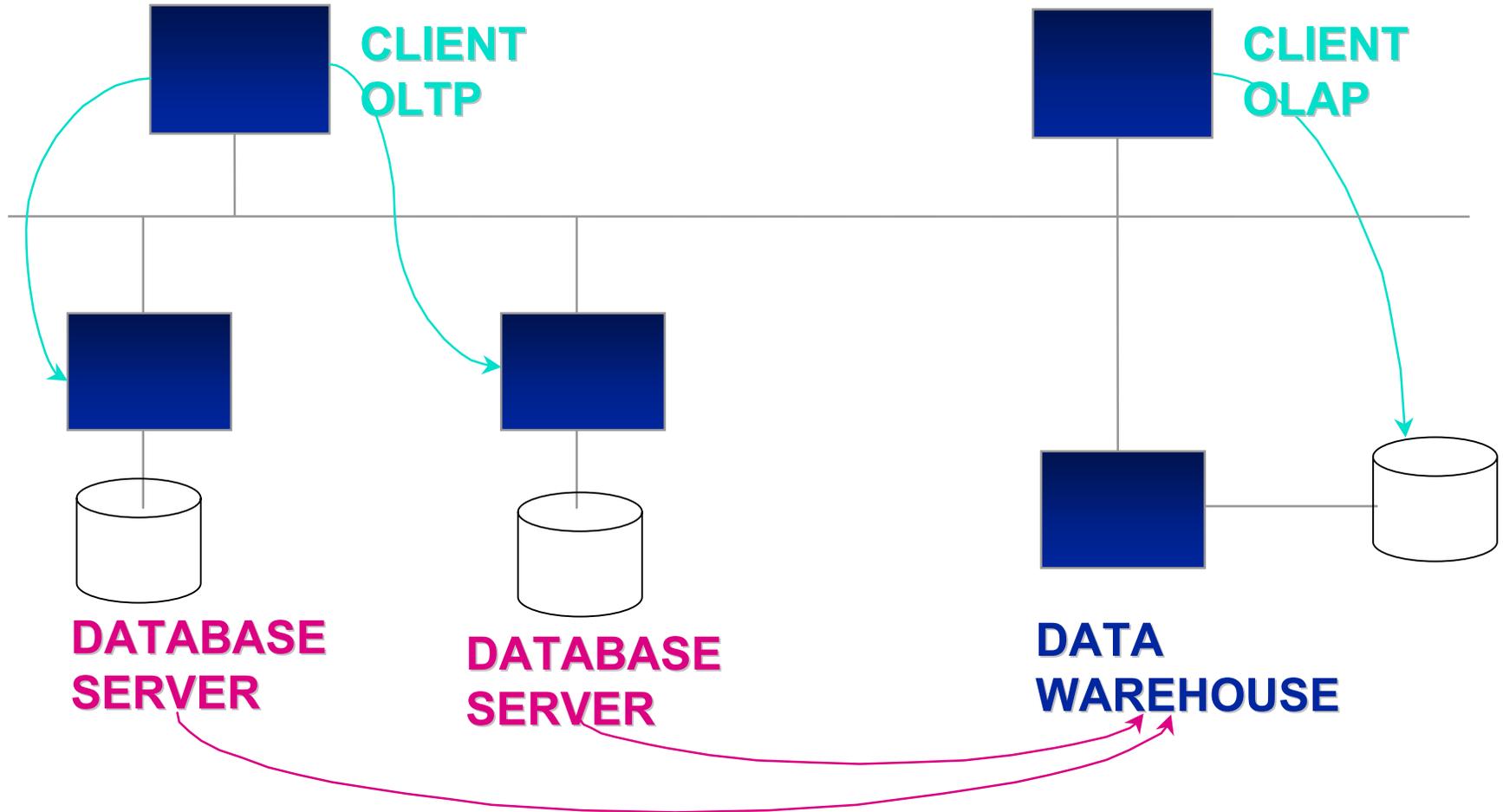
**Un caso particolare:
replica in computer mobili**

- **computer mobili :**
saltuariamente collegati ad una rete
- **copie disconnesse per ore o giorni
intere, poi riconnesse (riconciliazione)**
- **applicazione :**
agenti di vendita mobili

Separazione Funzionale degli Ambienti

- **ambiente operativo:**
gestione "in linea" dei dati,
finalizzato alla gestione delle modifiche
→ **On Line Transaction Processing (OLTP)**
- **ambiente di analisi:**
gestione "fuori linea" dei dati,
finalizzato ad interrogazioni complesse,
quali analisi statistiche, what-if...
→ **On Line Analytical Processing (OLAP)**

Ambiente per l'Analisi : Data Warehouse



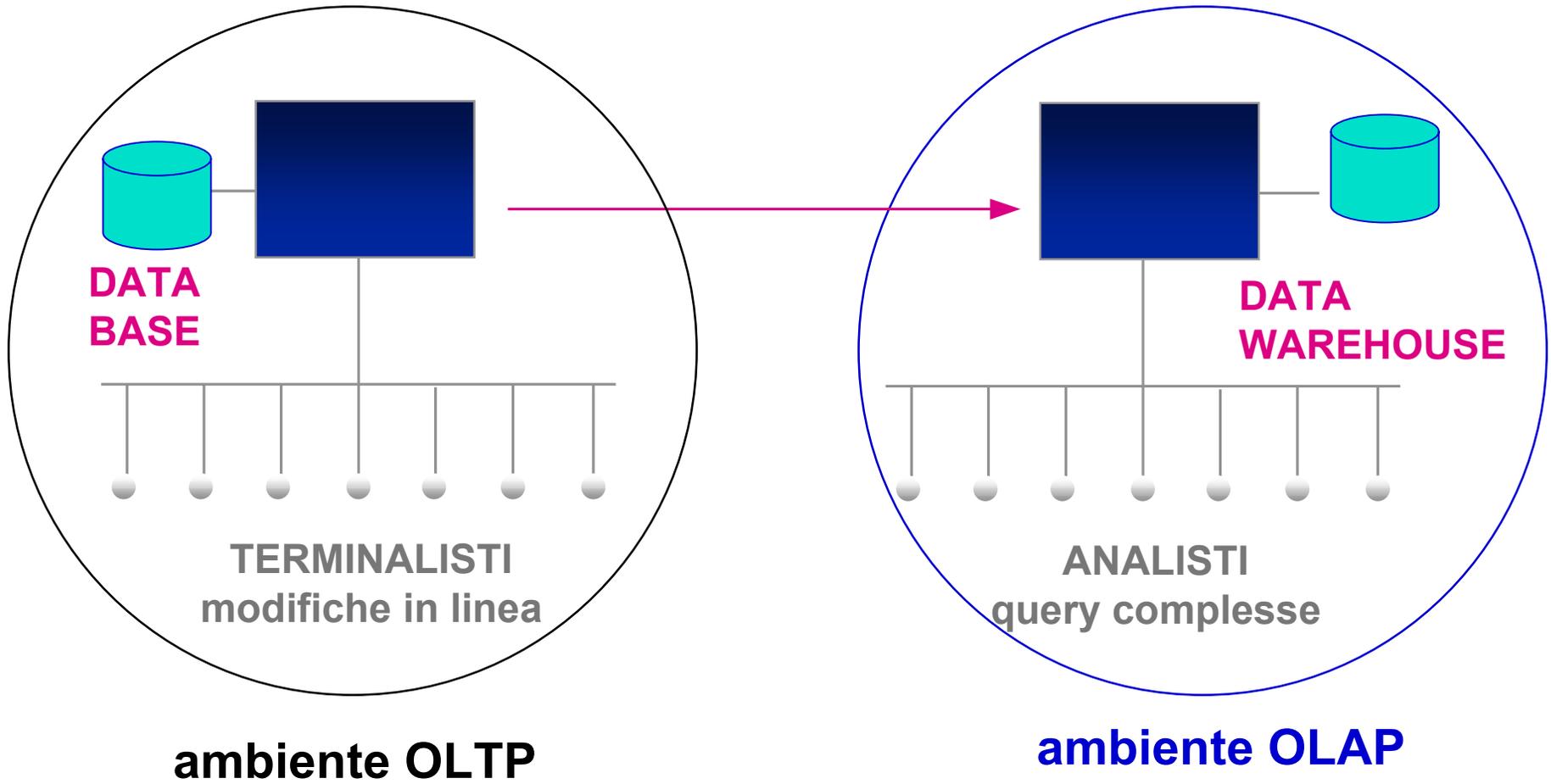
Data Warehouse

Definizione:

**insieme delle strutture dati
e degli strumenti necessari a ricavare
(partendo dai dati operazionali)
le informazioni necessarie per effettuare
una valutazione tecnico-economica
e un'analisi strategica
del comportamento dell'impresa**

- Analisi Multidimensionale**
- Data Mining**

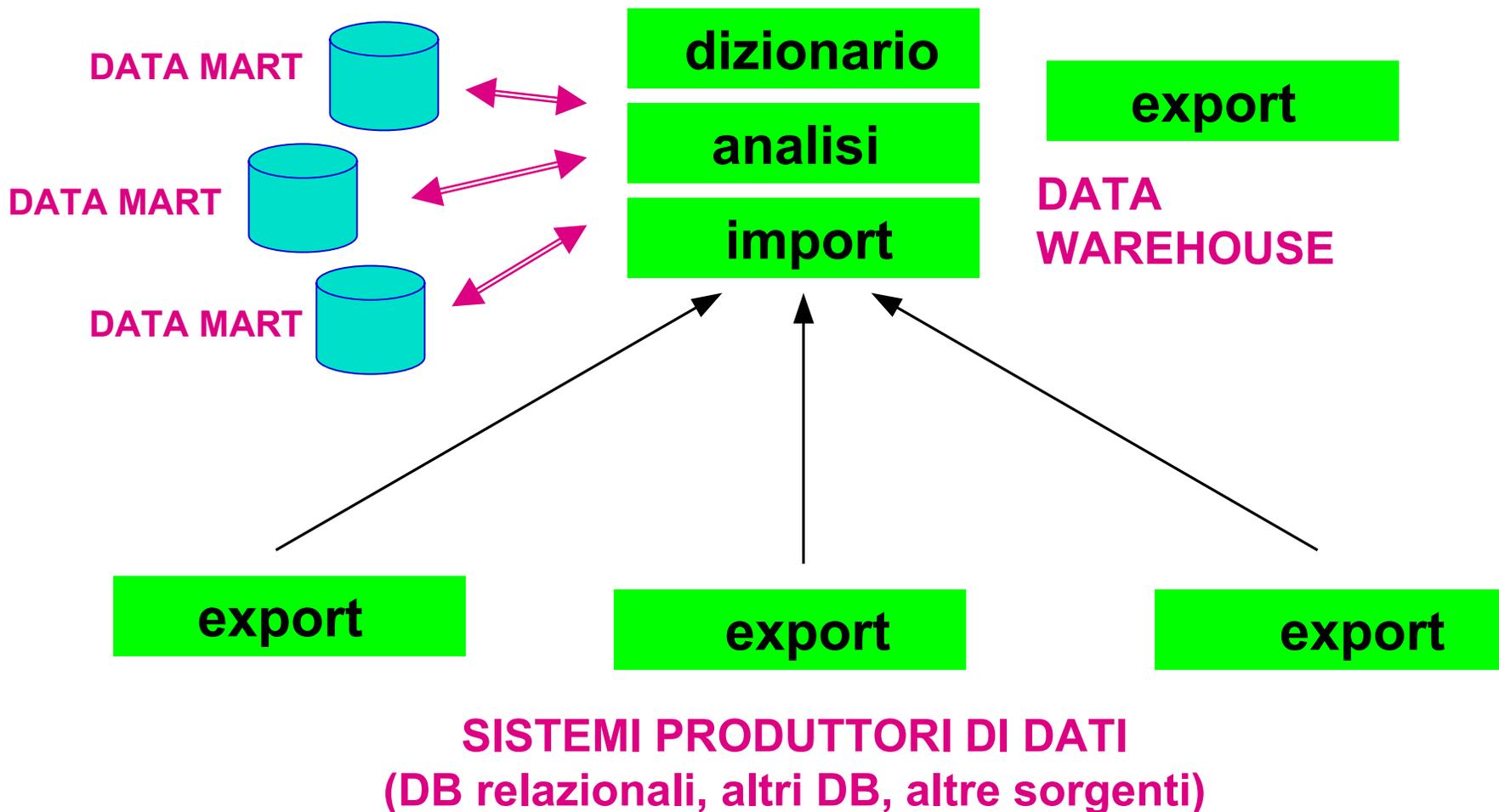
Separazione degli ambienti OLTP e OLAP



Motivazioni della separazione degli ambienti

- **tipologia di utenza differente**
- **organizzazione dei dati differente**
- **tecniche completamente innovative**
 - **tipo di query e di interazione**
 - **uso di strutture dati ad hoc e metodi di accesso dedicati**
 - **uso del parallelismo**

Moduli in un data warehouse



Differenze fra data warehouse e DBMS tradizionale

a **modalità d'uso**

- funzionamento "normale": query di sola lettura
- aggiornamento: lunghi programmi batch

b **progetto fisico dei dati**

- supporto accesso sequenziale
- "clusterizzazioni" e raggruppamenti predefiniti
- parallelismo intra-query

c **ambiente di sviluppo**

- orientati ad utenti finali non informatici

Problemi di progetto

- **qualità dei dati**
 - filtro di dati scorretti
 - integrazione di dati da fonti multiple ed eterogenee
- **modalità di acquisizione**
(basata su replication manager)
- **tecniche di analisi**