

Architetture Basi di dati

Architetture di Basi di dati distribuite, eterogenee, autonome

Fonti di riferimento

- Queste trasparenze parte 3
- Basi di Dati (Atzeni, Paraboschi, Torlone,)
- Tamer-Ozsu, P. Valduriez, *Principles of Distributed Database Systems*, 1999 Prentice Hall

Architettura di un DBMS centralizzato

Posti di lavoro



**Interrogazioni,
Aggiornamenti,
transazioni**

Applicazioni

Elaboratore

Amministratore della
Base dati

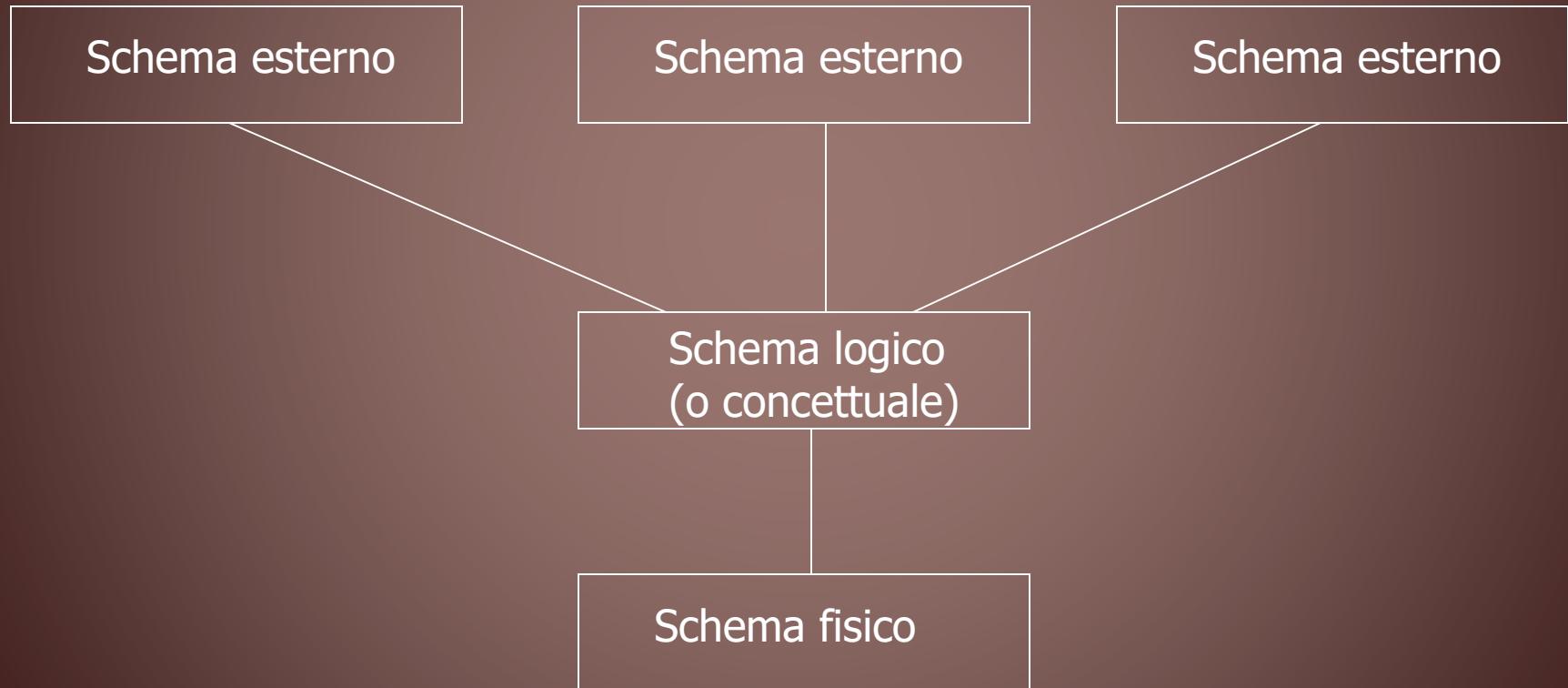
**Comandi
Data Description
Language**

DBMS

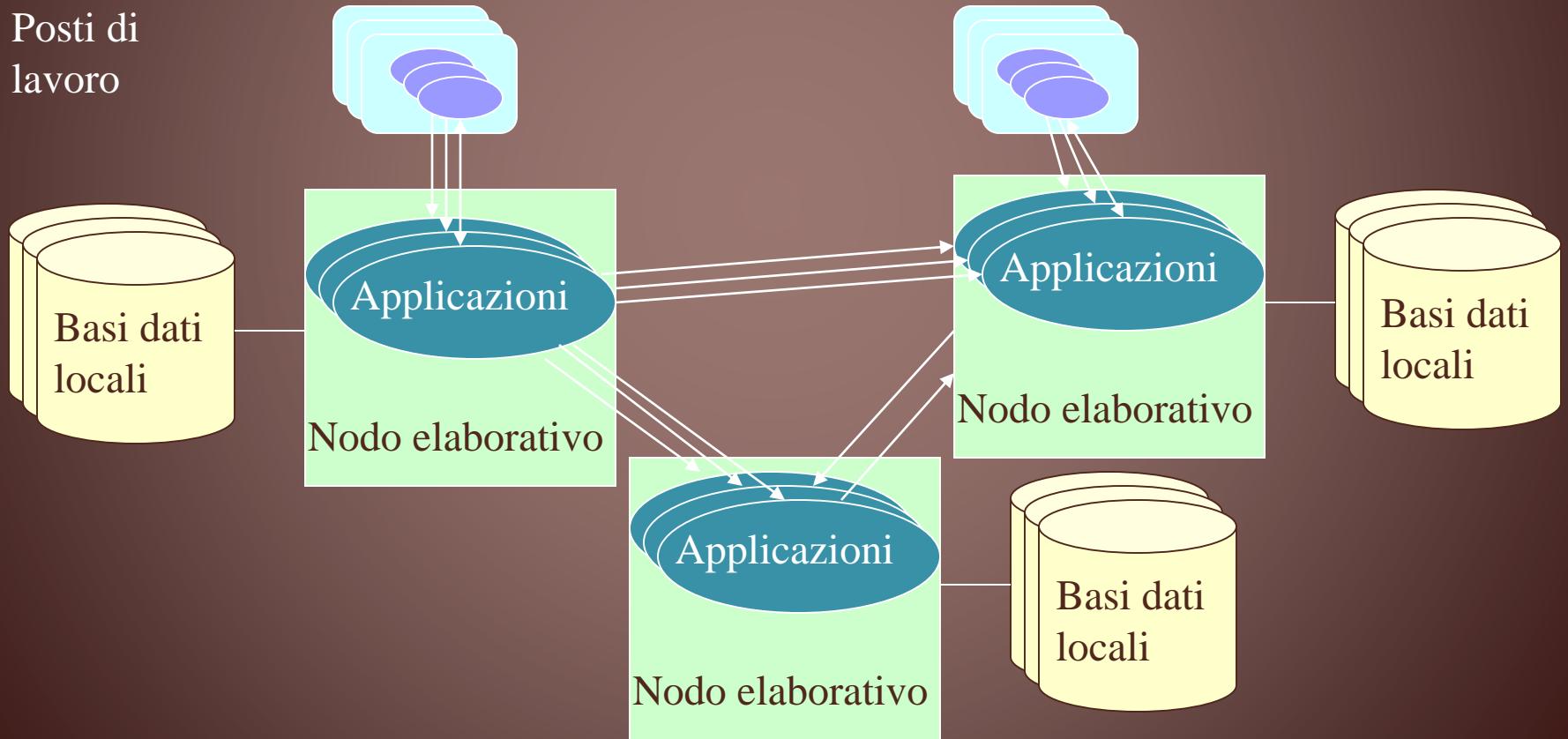


Basi di
dati

Architettura ANSI/SPARC



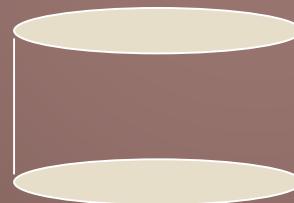
Architettura di un sistema DB distribuito



Esempio:
base dati distribuita

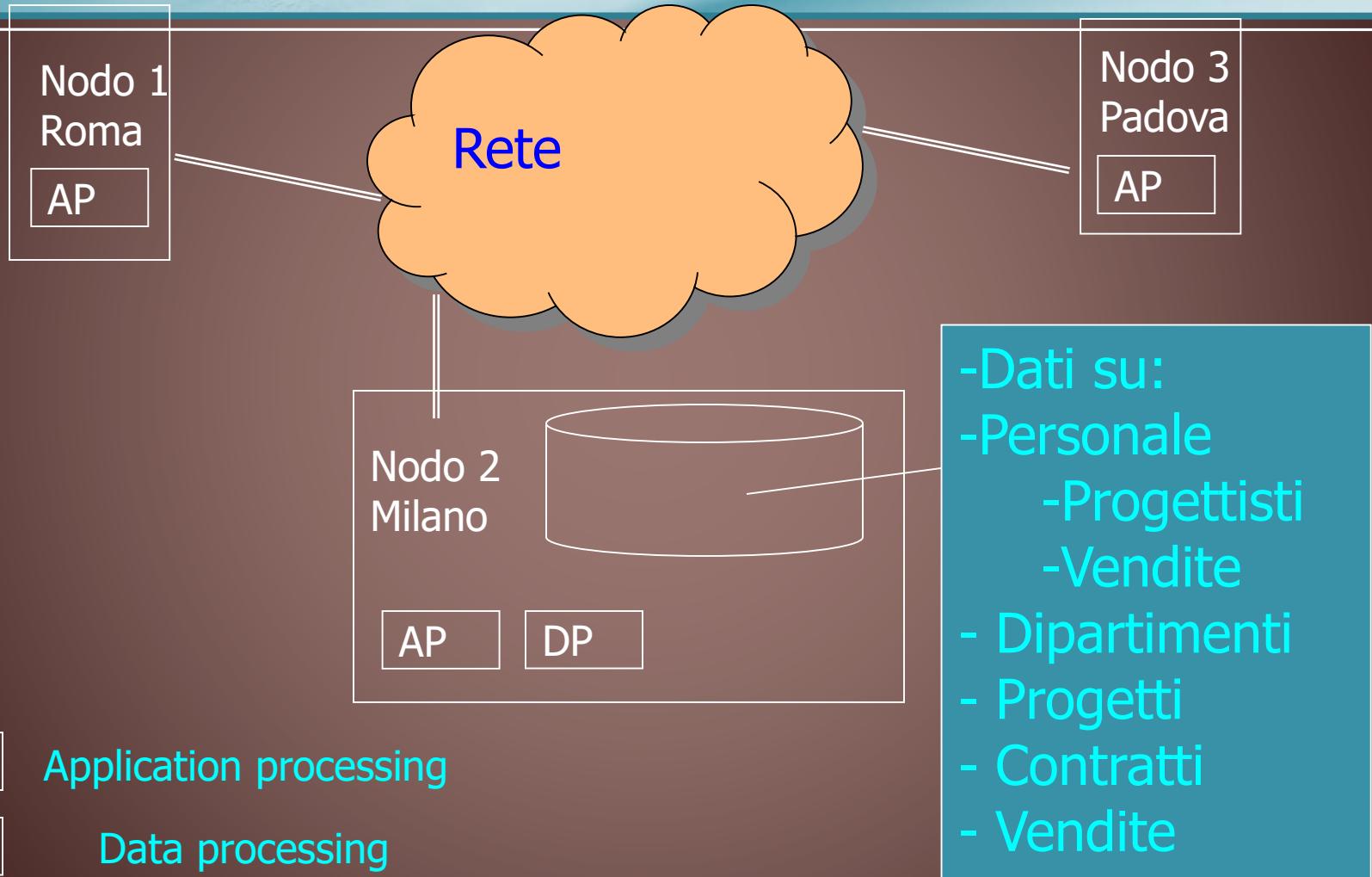
Passaggio, in una singola
organizzazione, da base dati
centralizzata a distribuita

Struttura della organizzazione

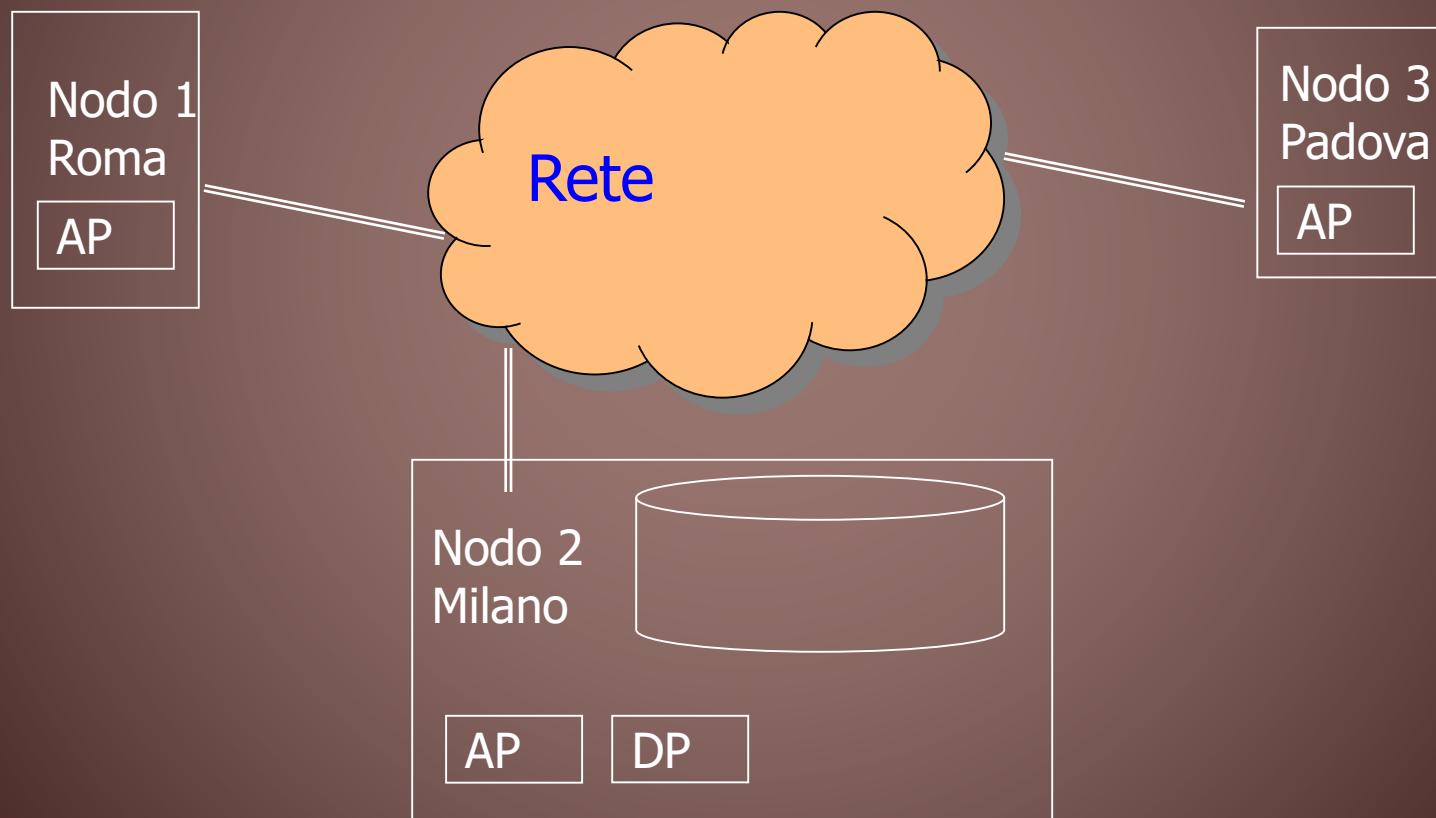


Unica base di dati,
localizzata a Milano
acceduta da tutti i nodi con
transazioni centralizzate

Base Dati centralizzata

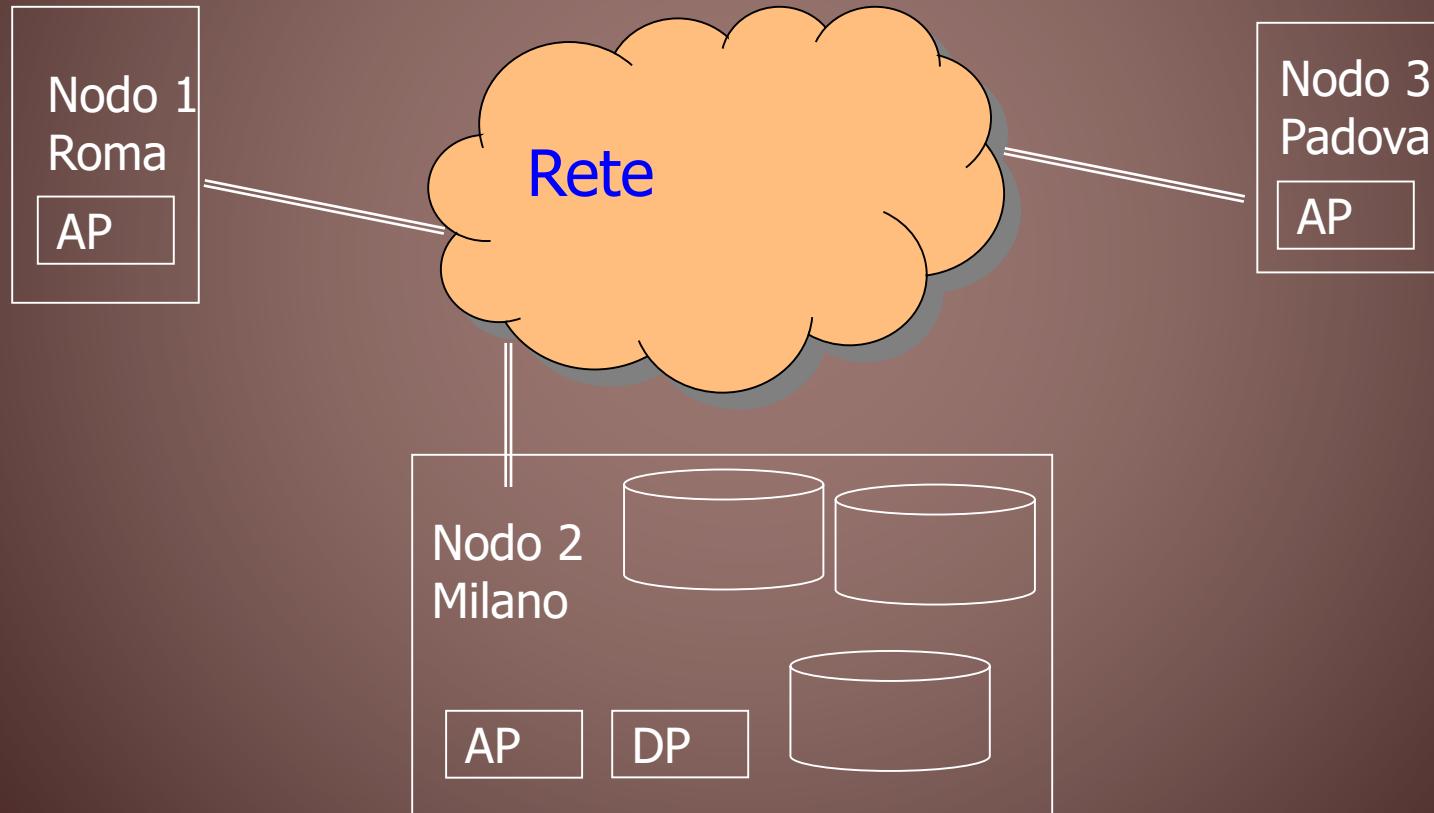


Da base dati centralizzata a base dati distribuita



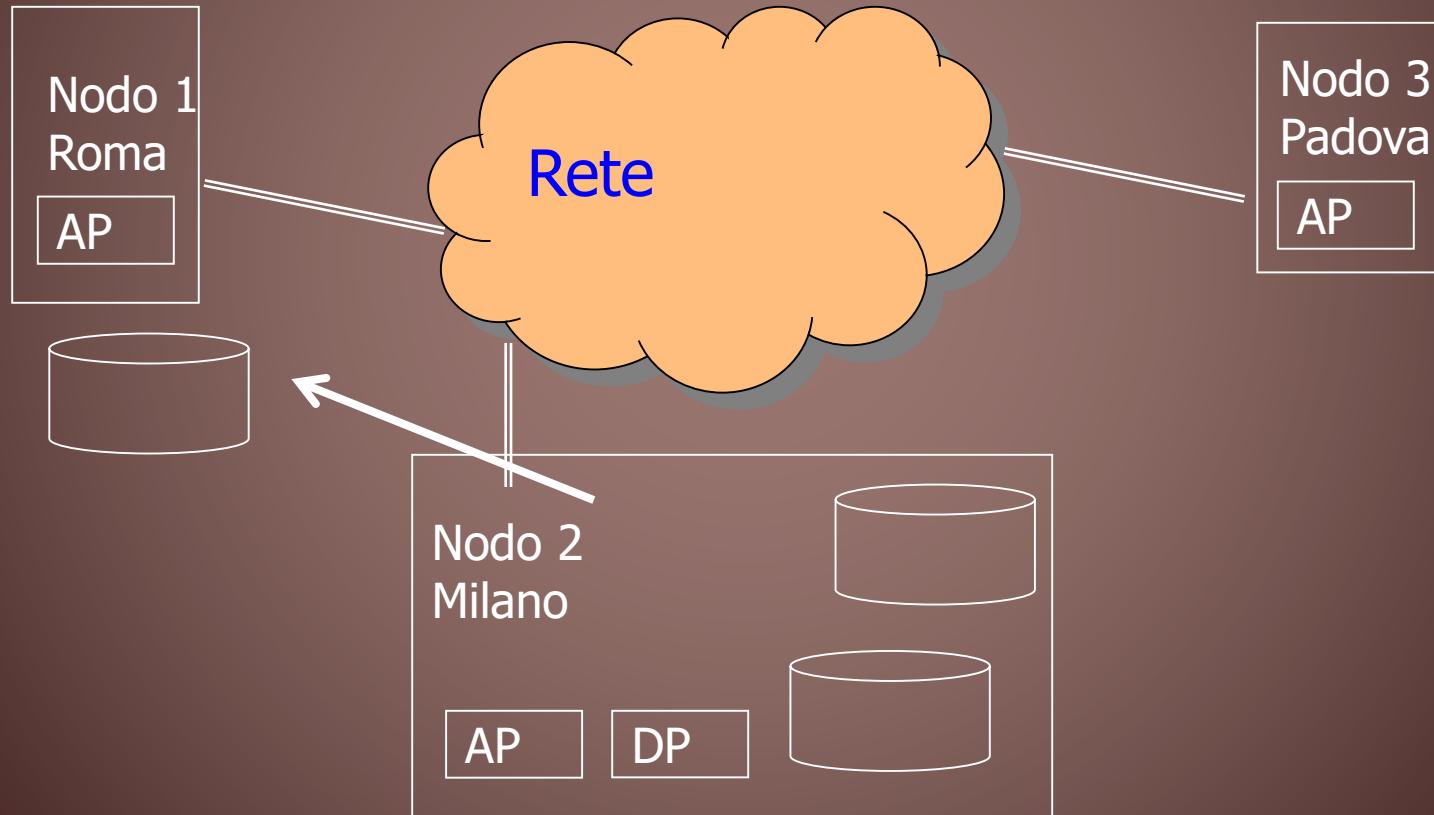
Può essere utile per esigenze di efficienza o affidabilità spostare o duplicare alcuni dati presso altri nodi

Da base dati centralizzata a base dati distribuita



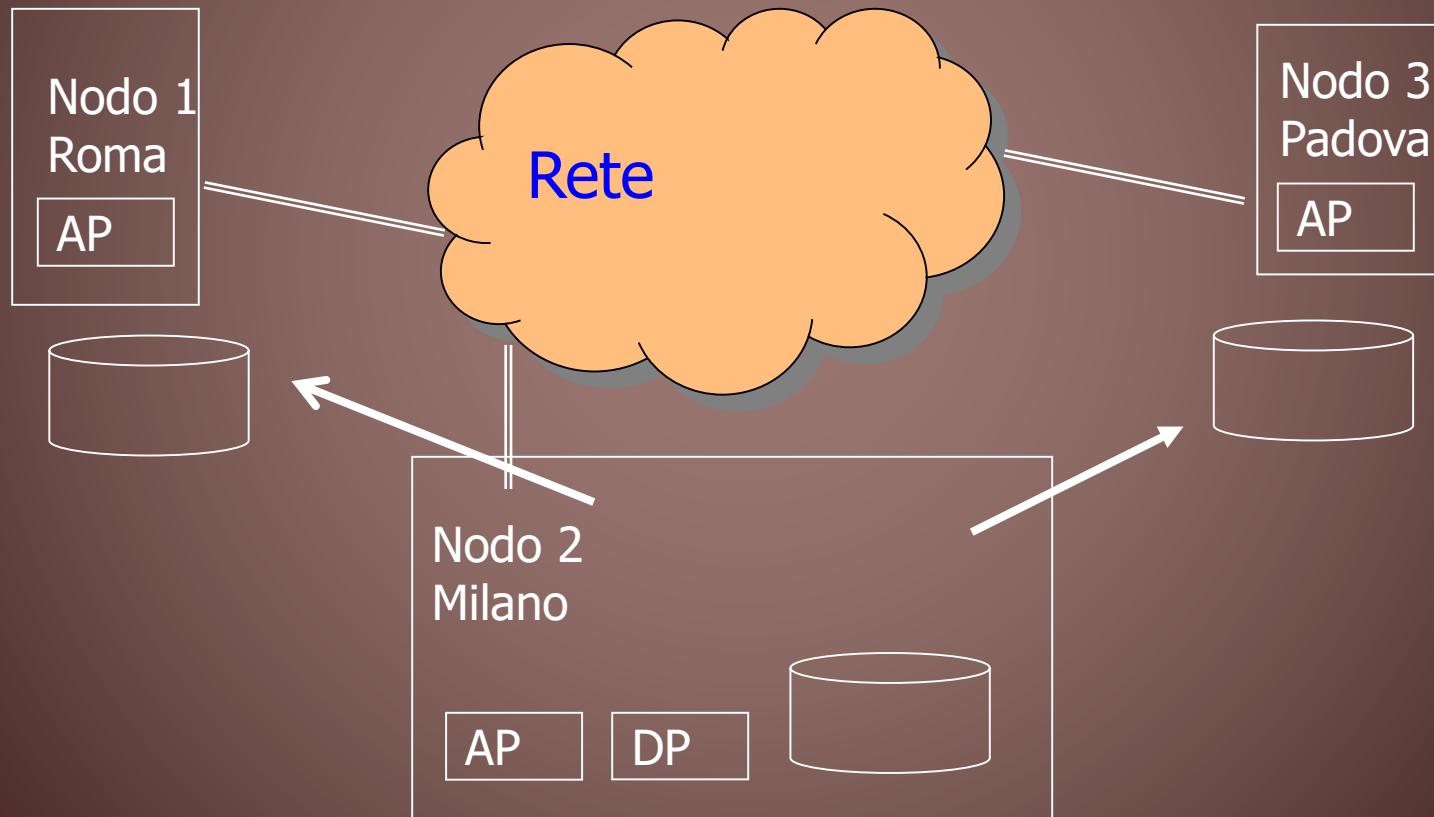
.... frammentando la base di dati in frammenti, scelti ad es.
sulla base degli accessi dai vari nodi

Da base dati centralizzata a base dati distribuita



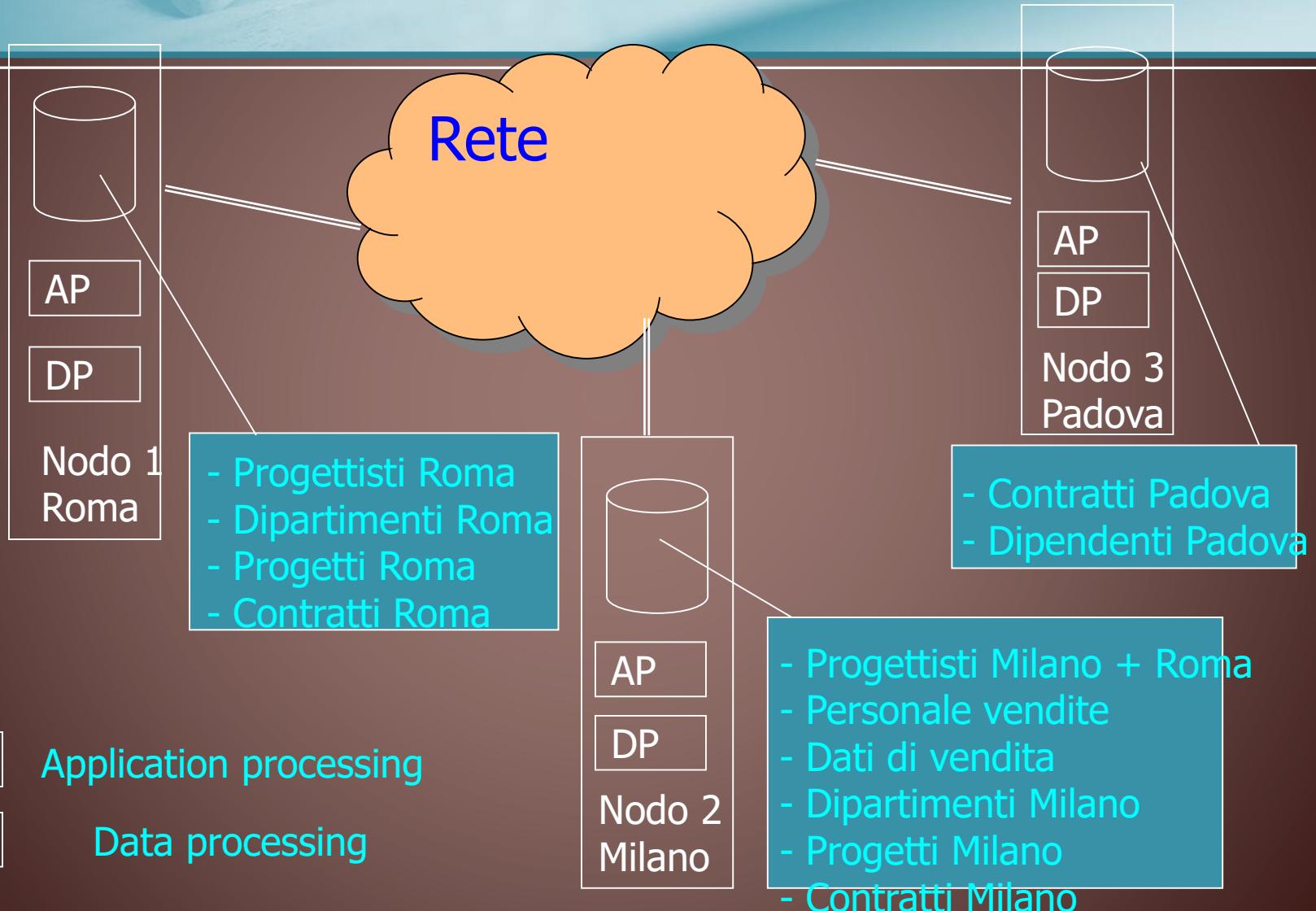
E spostando i frammenti nei vari nodi ...

Da base dati centralizzata a base dati distribuita

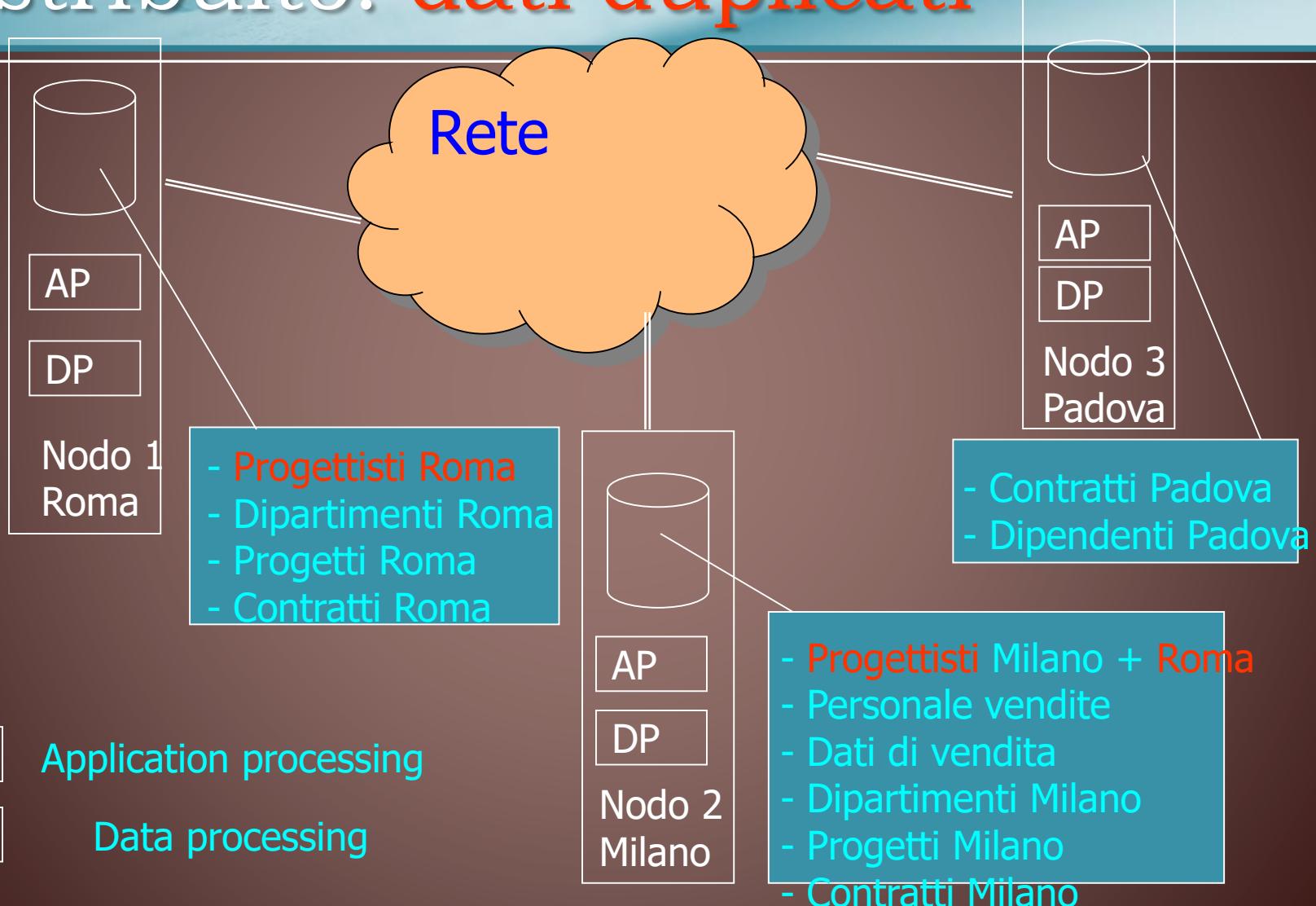


E spostando i frammenti nei vari nodi ...

Base dati distribuita con DBMS distribuito



Base dati distribuita con DBMS distribuito: dati duplicati



Confronto tra tipologie di sistemi

Sistema	Distribuzione	Eterogeneita'	Autonomia
Base dati distribuita			
Multidatabase (tematica informazione)			

Confronto tra tipologie di sistemi

Sistema	Distribuzione	Eterogeneita'	Autonomia
Base dati distribuita	Alta		
Multidatabase	Alta		

Confronto tra tipologie di sistemi

Sistema	Distribuzione	Eterogeneita'	Autonomia
Base dati distribuita	Alta	Bassa	
Multidatabase	Alta	Alta	

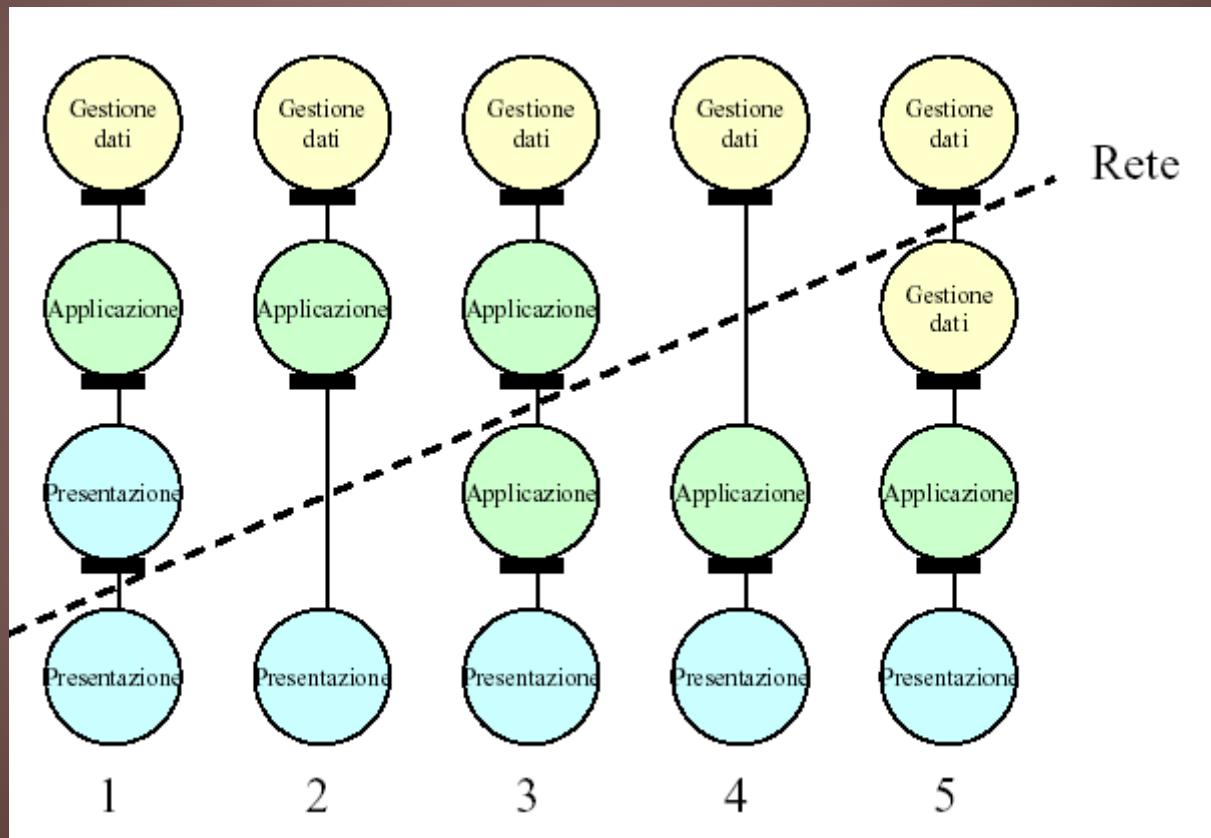
Confronto tra tipologie di sistemi

Sistema	Distribuzione	Eterogeneita'	Autonomia
Base dati distribuita	Alta	Bassa	Bassa
Multidatabase	Alta	Alta	Alta

Distribuzione

- Nella realizzazione di un sistema con caratteristiche di distribuzione si pone il **problema di cosa distribuire.**
- Sistema distribuito sistemi per cui si verifica almeno una delle seguenti condizioni:
 - Le applicazioni, fra loro cooperanti, risiedono su più nodi elaborativi (elaborazione distribuita)
 - L'archivio informativo è distribuito su più nodi (base di dati distribuita)
- Ad esempio la *workstation* ed un *server* possono ripartirsi le funzionalità di elaborazione, oppure su una schiera di sistemi server possono essere ripartiti diversi sottoprogrammi che implementano diverse funzionalità

Distribuzione



Distribuzione

- La distribuzione in un livello è ortogonale e trasparente agli altri.
- È importante in fase di definizione architettonica capire cosa distribuire
- Domanda
 - Cosa distribuire nei seguenti casi
 - Gestione studenti scuole superiori
 - Gestione studenti universitari
 - Sito Web
 - Portale Web

Distribuzione

- Considerare l'esistente
- Nei processi di merger & acquisition due organizzazioni si fondono (o una compra l'altra)
- Problema tecnologico il nuovo sistema informativo deve
- Unire due le due logiche applicative
- Unire i due sistemi di archiviazione
- Altro?



Criteri di classificazione dei Distributed DB

Modelli architetturali per DDBMS -

1

- Un DBMS Distribuito Eterogeneo Autonomo e' in generale una **federazione** di DBMS che collaborano nel fornire servizi di accesso ai dati con livelli di **trasparenza** definiti
- Per **trasparenza** intendiamo la proprietà generale di nascondere le diversità tra basi di dati nei nodi del sistema, negli aspetti di **distribuzione, eterogeneità, autonomia**.

Spinta alla creazione di basi di dati e sistemi informativi federati

- L'esigenza di **integrare a posteriori** (o talvolta **progettare fin dall'inizio**) sistemi di tipo federato emerge in molti casi:
 - Evoluzione e/o integrazione di **componenti applicativi** sviluppati separatamente (per ragioni tecniche, organizzative o temporali)
 - cooperazione di **processi** in precedenza separati
 - cooperazione (o fusione) di **enti** o **aziende** indipendenti
- ed è stimolata dallo **sviluppo delle reti**

Modelli architetturali per DDBMS

- E' possibile distinguere diverse modalita' di federazione e diverse corrispondenti architetture
- Dimensioni *ortogonali* utili per la classificazione
 - Autonomia
 - Distribuzione
 - Eterogeneita'

Autonomia

Fa riferimento al grado di indipendenza tra i nodi

Si distinguono diverse forme di autonomia

- **Di progetto:** ogni nodo adotta un proprio modello dei dati e sistema di gestione delle transazioni
- **Di condivisione:** ogni nodo sceglie la porzione di dati che intende condividere con altri nodi
- **Di esecuzione:** ogni nodo decide in che modo eseguire le transazioni che gli vengono sottoposte

Tipi di autonomia

- DBMS Strettamente integrati (nessuna autonomia)
 - Dati *logicamente* centralizzati
 - Un unico data manager responsabile delle transazioni applicative
 - I data manager locali non operano in modo autonomo
- Semi-autonomi
 - Ogni data manager e' autonomo ma partecipa a transazioni globali
 - Una parte dei dati e' condivisa
 - Richiedono modifiche architetturali per poter fare parte della federazione
- Totalmente autonomi (o Peer to Peer)
 - Ogni DBMS lavora in completa autonomia ed e' inconsapevole dell'esistenza degli altri

Distribuzione

- Fa riferimento alla distribuzione dei dati. Si possono distinguere:
- **Distribuzione client/server**, in cui la gestione dei dati e' concentrata nei server, mentre i client forniscono l'ambiente applicativo e la presentazione.
- **Distribuzione peer-to-peer**, in cui non c'e' distinzione tra client e server, e tutti i nodi del sistema hanno identiche funzionalita' DBMS
- **Nessuna distribuzione**
- Nel seguito non distingueremo tra distribuzione client/server e peer-to-peer

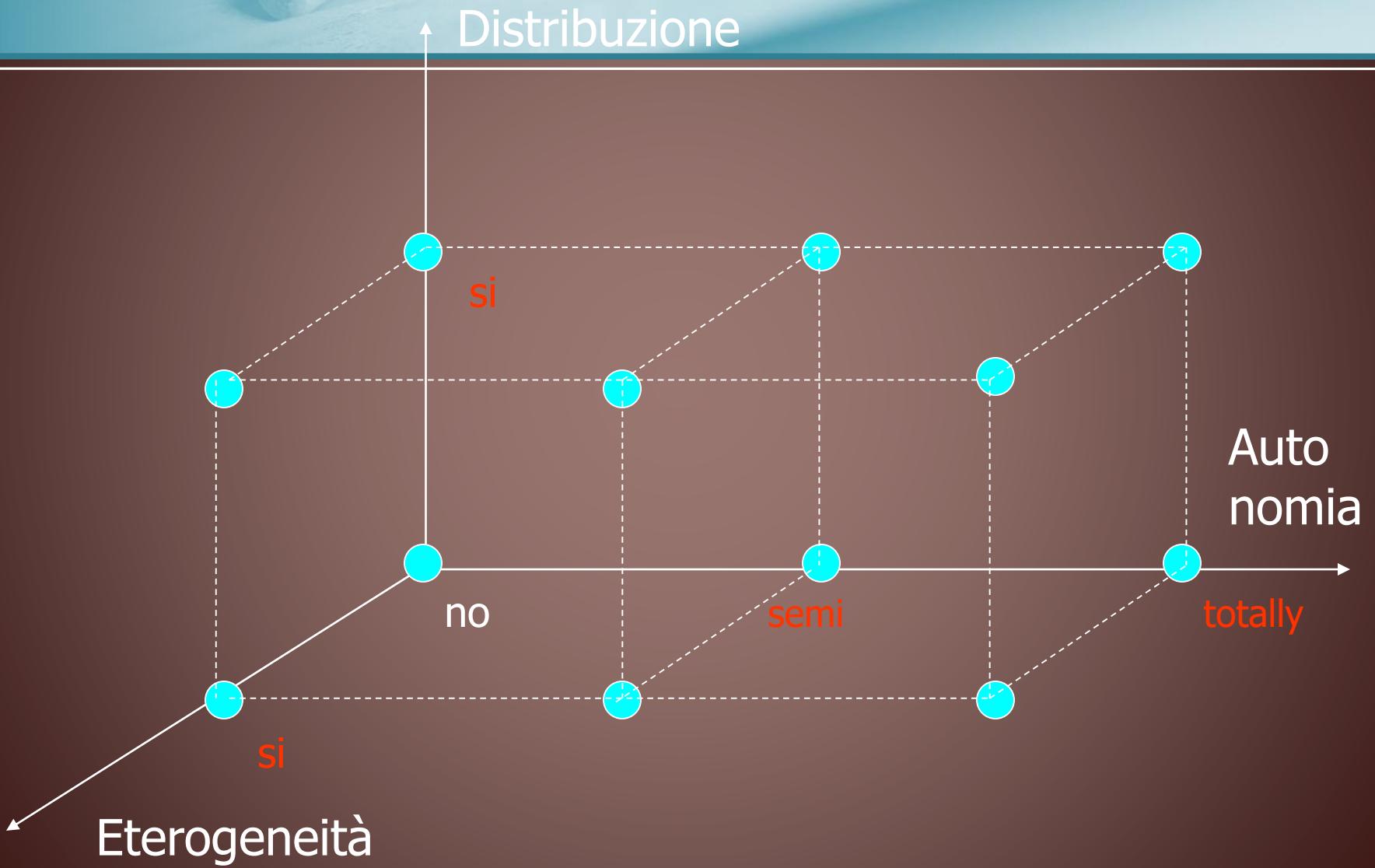
Eterogeneita'

- Puo' riguardare
- **Modello dei dati**
 - Es. Relazionale, XML, Object Oriented
- **Linguaggio di query**
 - Es. Diversi dialetti SQL, Query by Example, Linguaggi di interrogazione OO o XML
- **Gestione delle transazioni**
 - Es. Diversi protocolli per il concurrency control e per il recovery
- **Schema concettuale e schema logico**
 - Es. Un concetto rappresentato in uno schema come attributo e in un altro come entita'

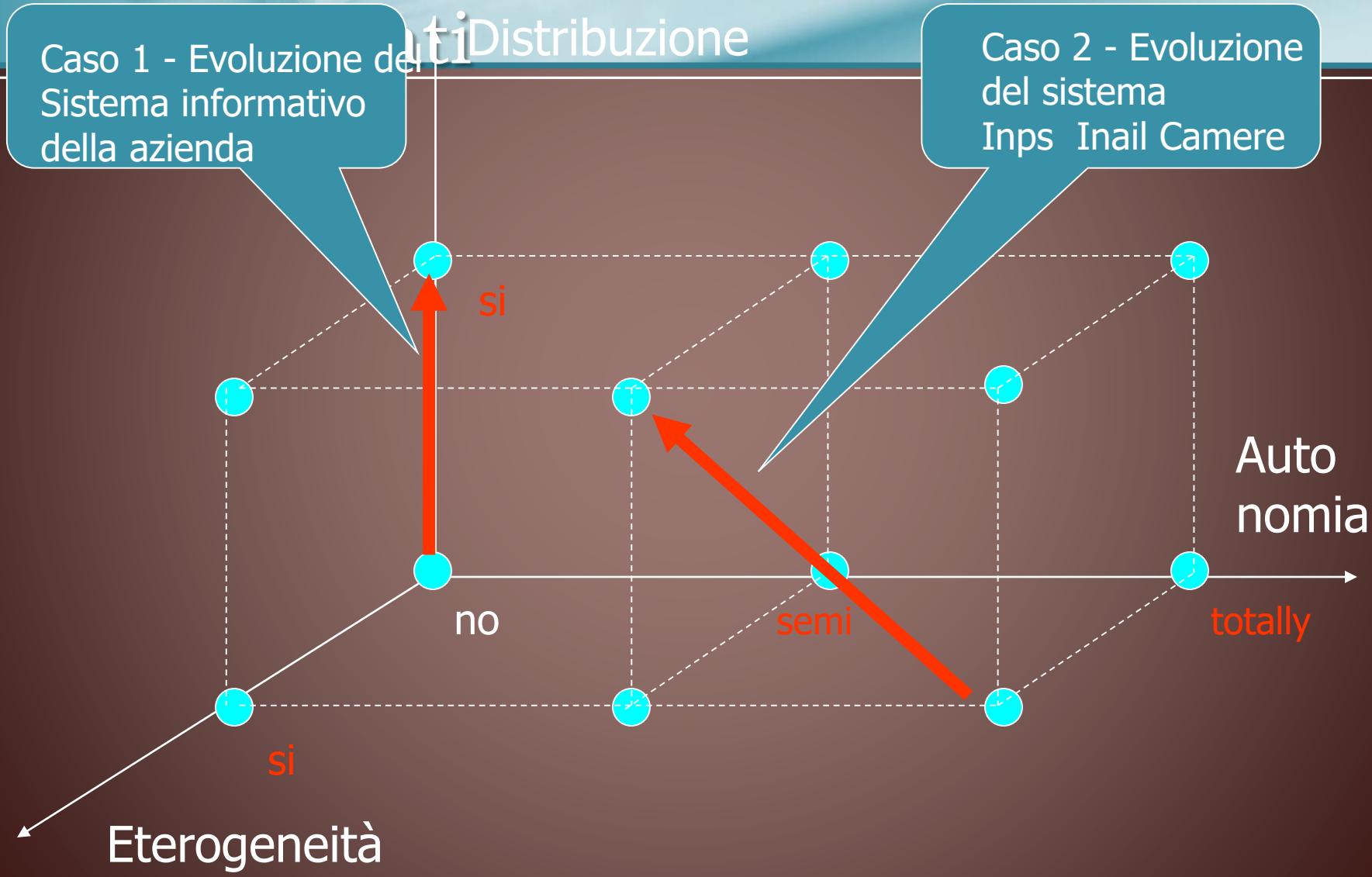


Tipologie di sistemi di BD DEA

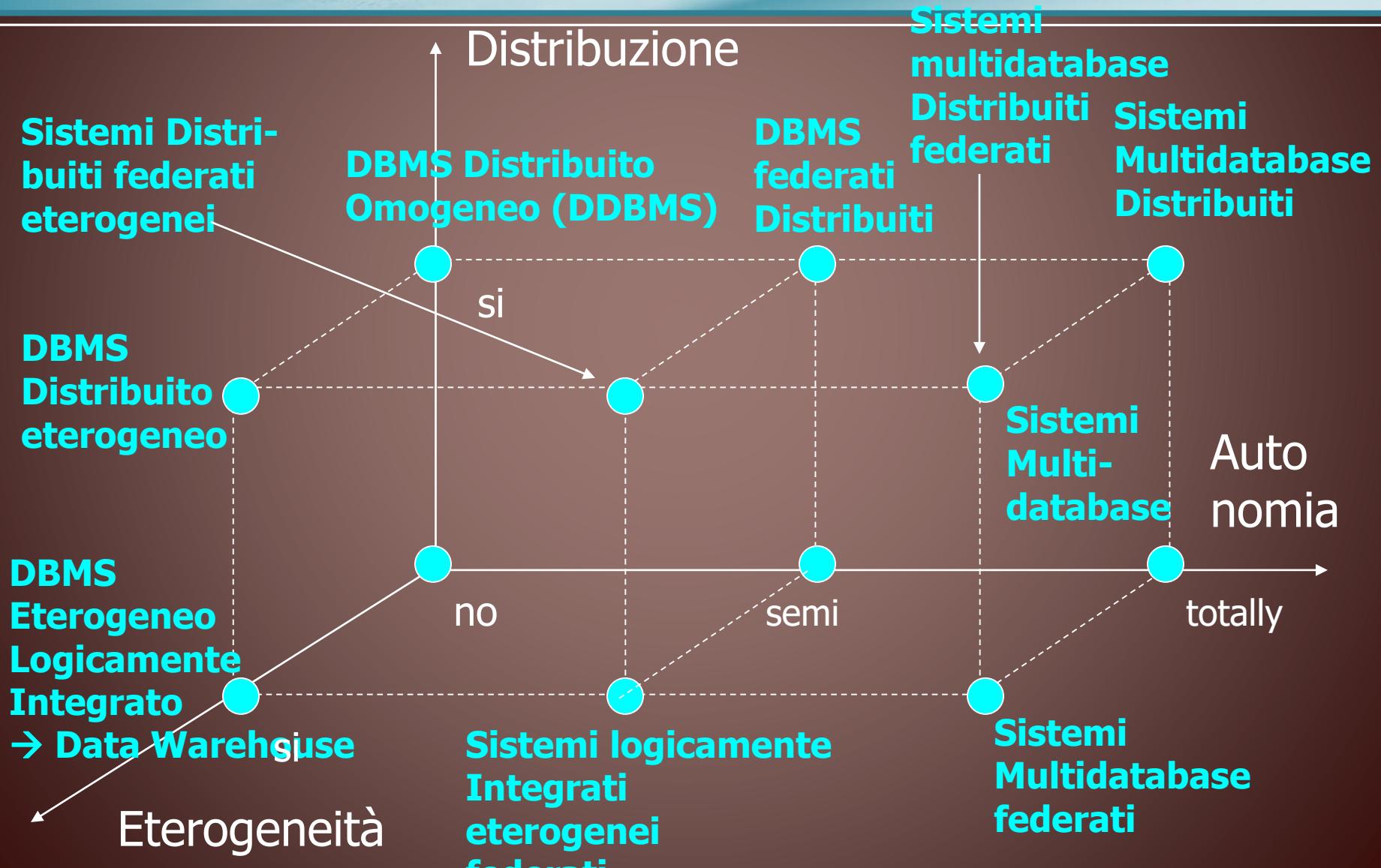
Tipologie di sistemi di BD distribuite



distribuite e collocazione dei due casi



Tipologie di sistemi di BD distribuite: tipi di sistemi



Tipologie di sistemi di BD distribuite tipologie rilevanti

Distribuzione

Sistemi Distri-
buiti federati
eterogenei

DBMS Distribuito
Omogeneo (DDBMS)

DBMS
federati
Distribuiti

Sistemi
multidatabase
Distribuiti

Sistemi
Multidatabase
Distribuiti

DBMS
Distribuito
eterogeneo

DBMS
logicamente
Integrato

DBMS
Eterogeneo
Logicamente
Integrato
→ DW

Sistemi
logicamente
Integrati
eterogenei
federati

Eterogeneità

Si

no

semi

si

Auto
nomia

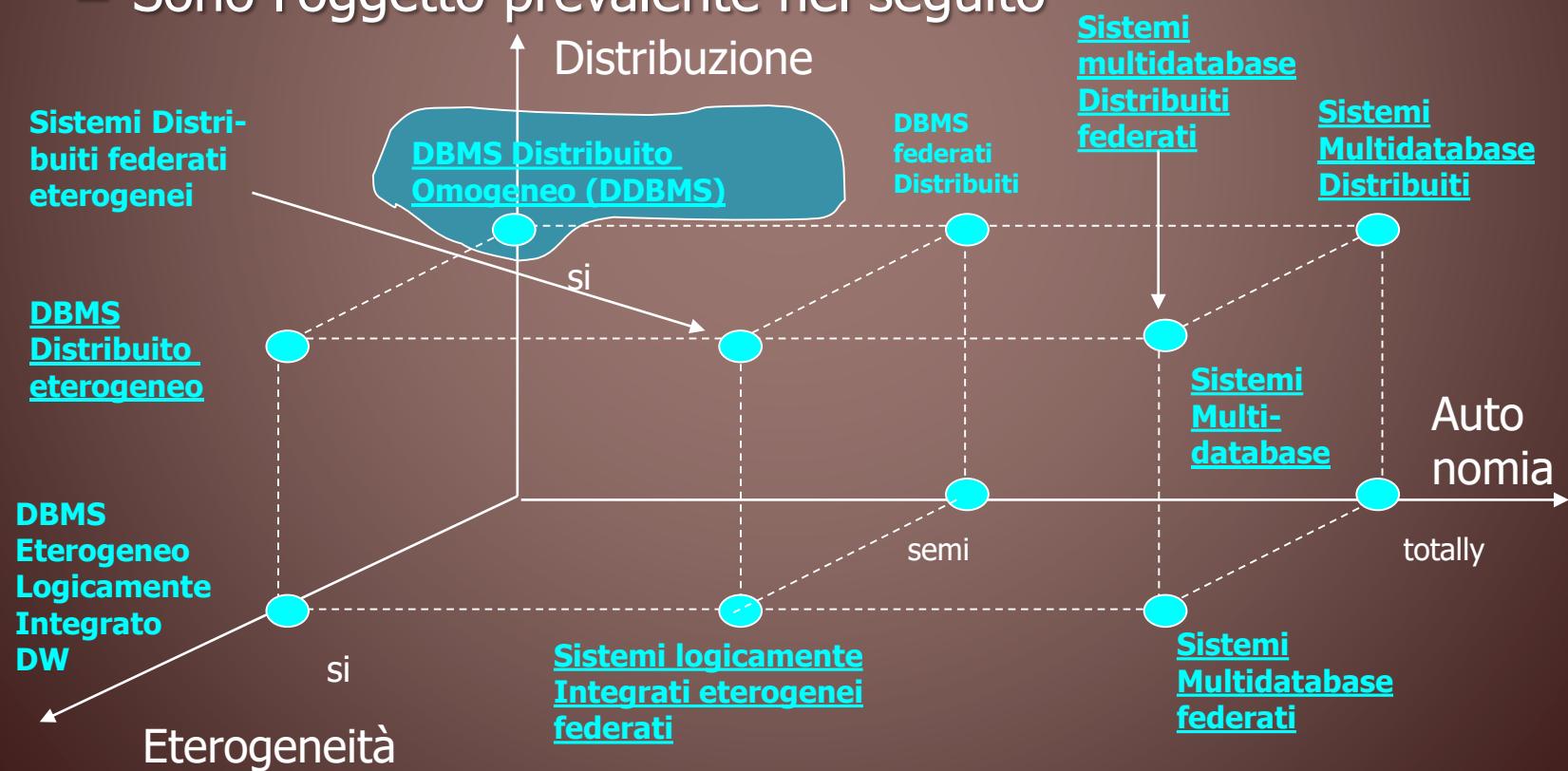
totally

Sistemi
Multi-
database

Sistemi
Multidatabase
federati

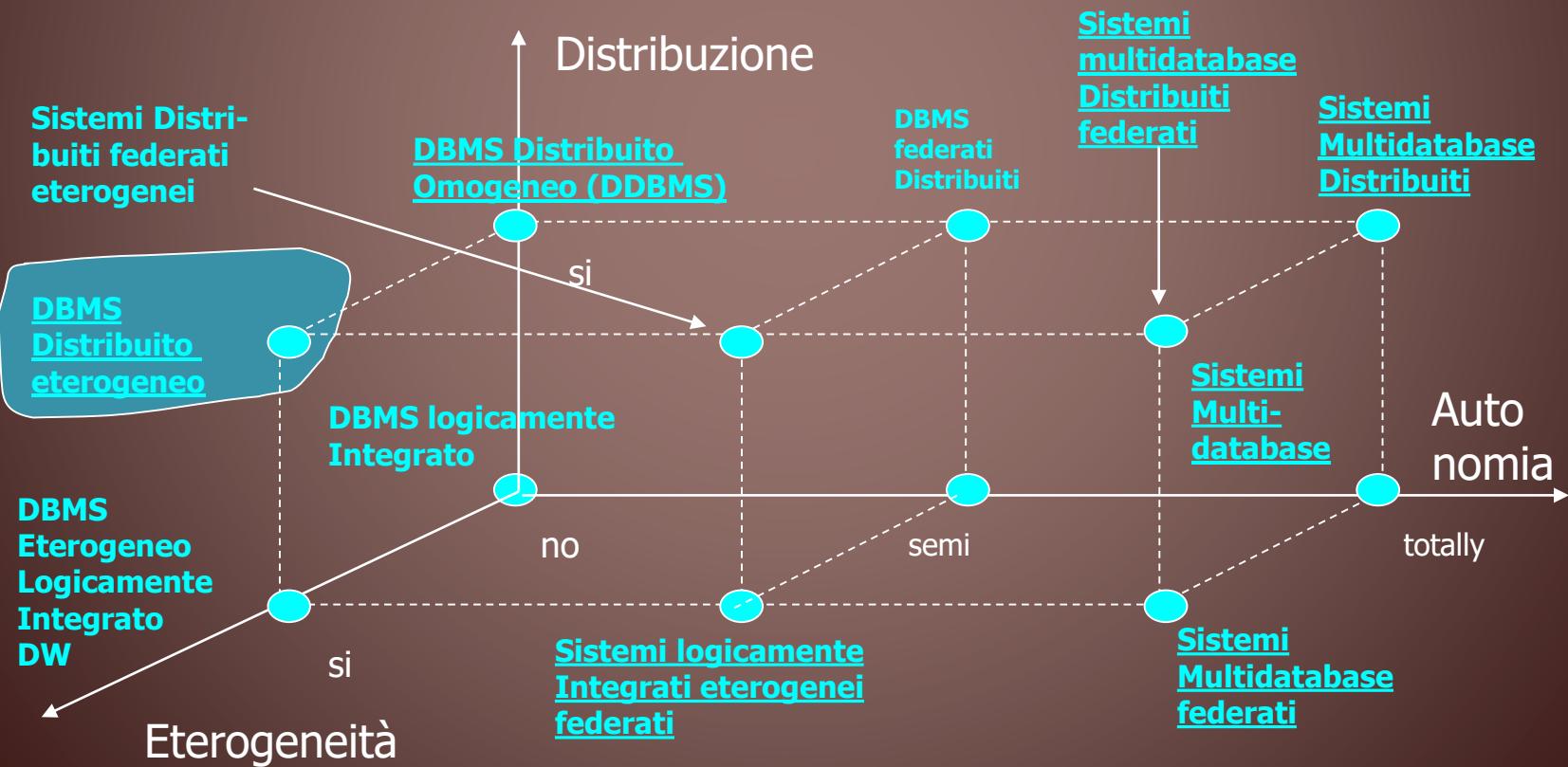
Tipologie piu' rilevanti - 1

- DBMS distribuiti omogenei (DDBMS): strettamente integrati, omogenei
 - Sono l'oggetto prevalente nel seguito



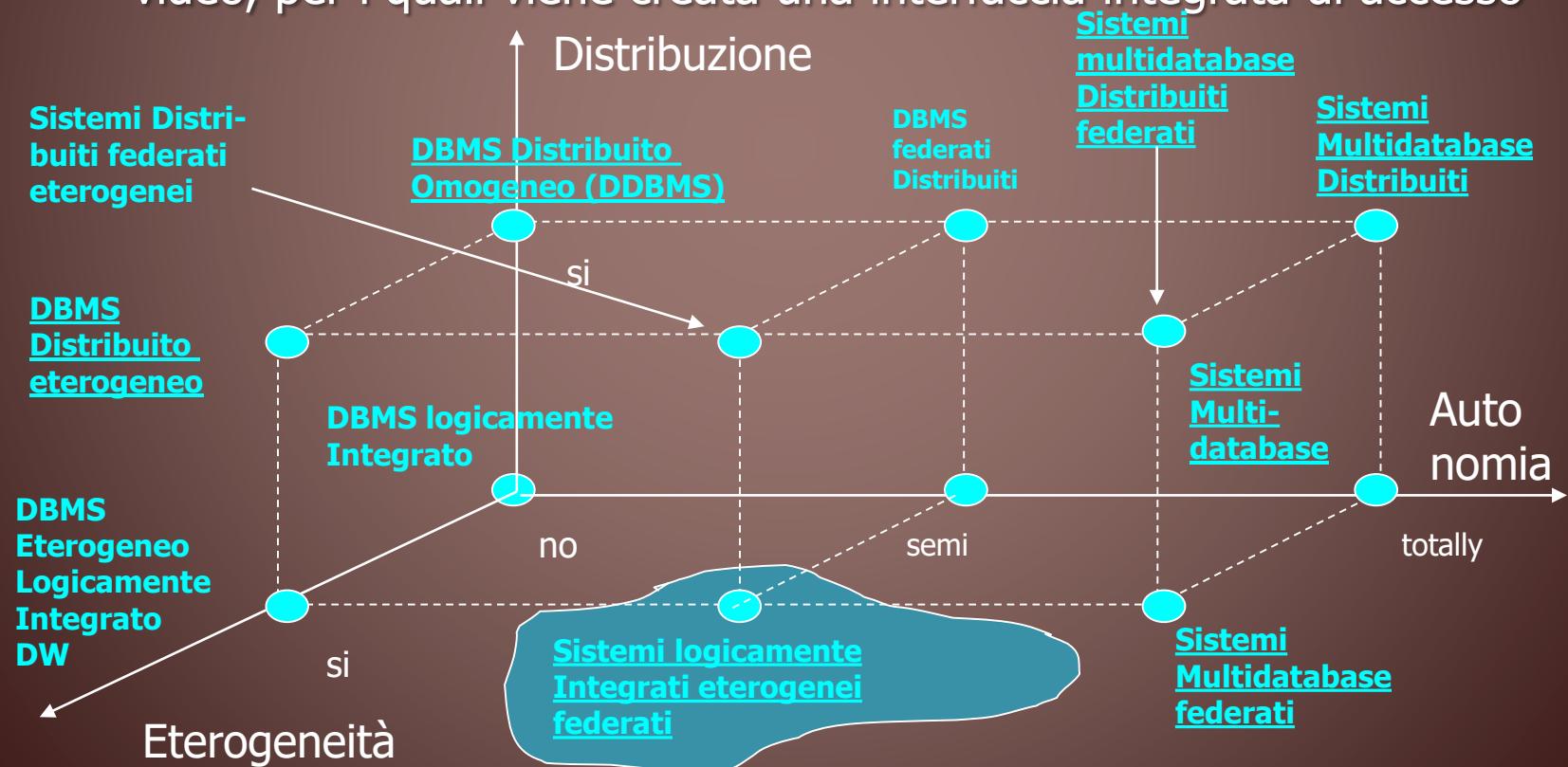
Tipologie piu' rilevanti - 2

- DBMS distribuiti eterogenei: strettamente integrati ma con eterogeneita' di varia natura



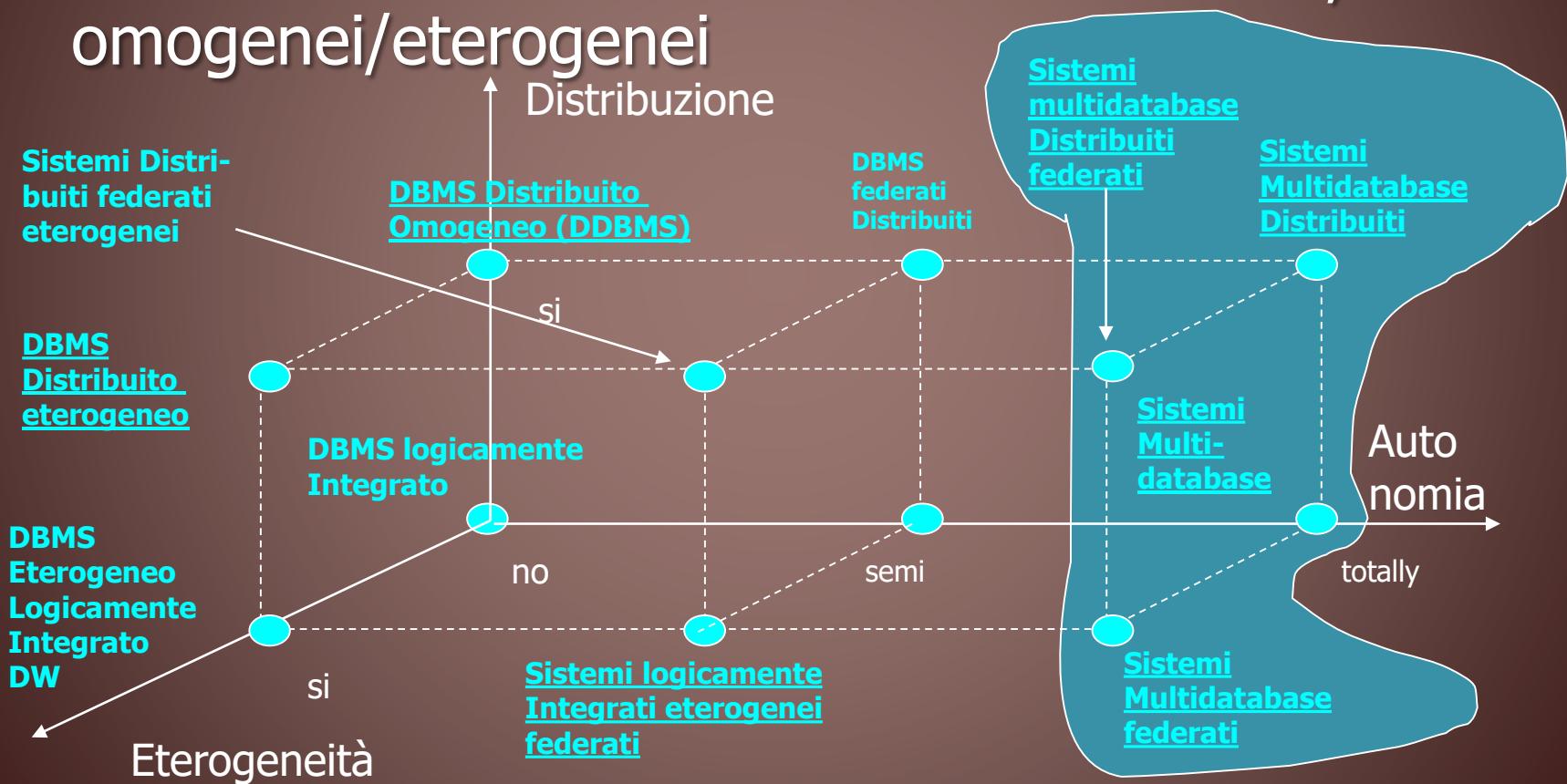
Tipologie piu' rilevanti - 3

- DBMS federati e logicamente integrati: semi-autonomi, eterogenei
 - Esempio: insieme di DBMS che gestiscono dati strutturati, immagini, video, per i quali viene creata una interfaccia integrata di accesso



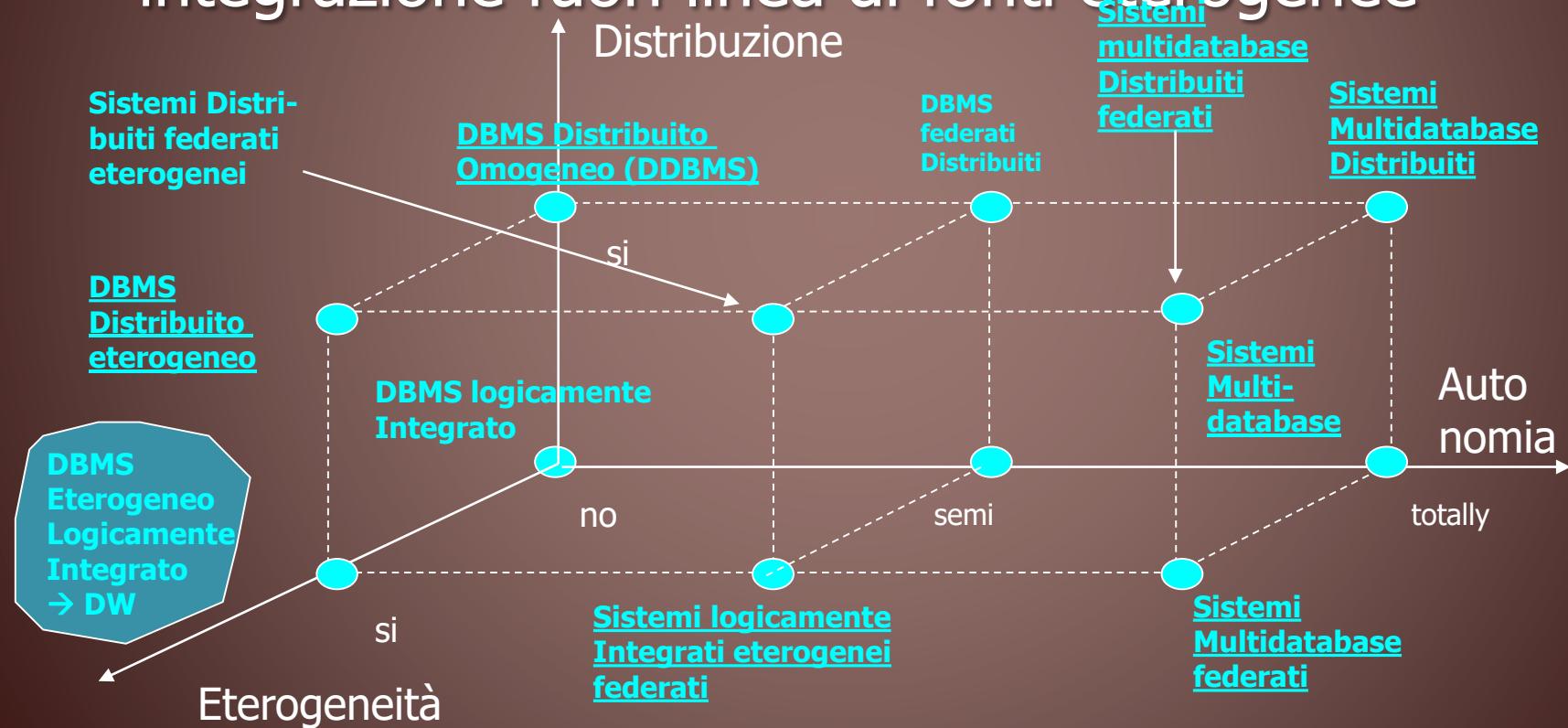
Tipologie piu' rilevanti - 4

- Multi Data Base MS: totalmente autonomi, omogenei/eterogenei



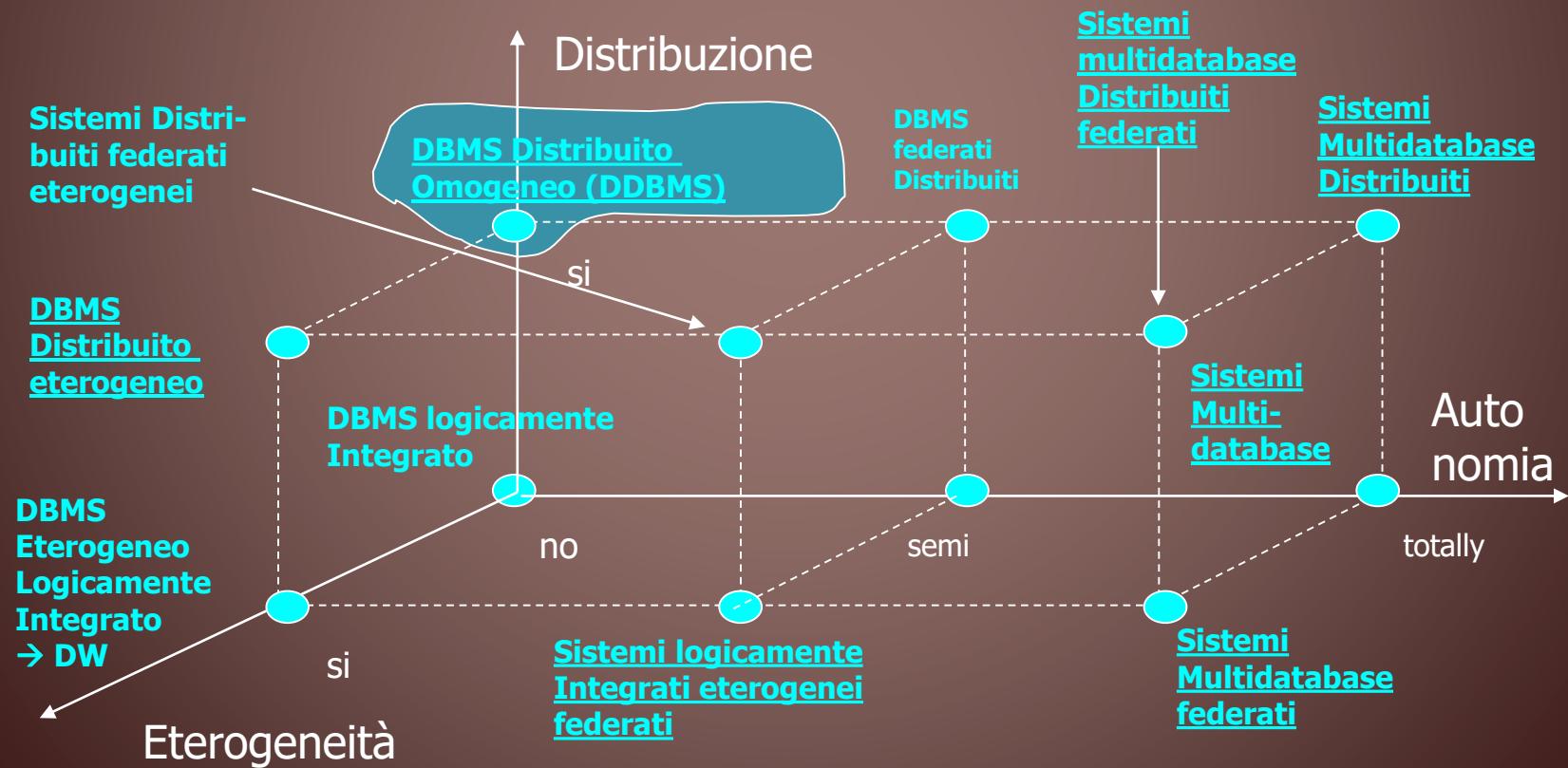
Tipologie piu' rilevanti - 5

- **Data Warehouse:** centralizzati, ma risultato della integrazione fuori linea di fonti eterogenee



Nel seguito ci concentreremo:

- DBMS distribuiti (omogenei)



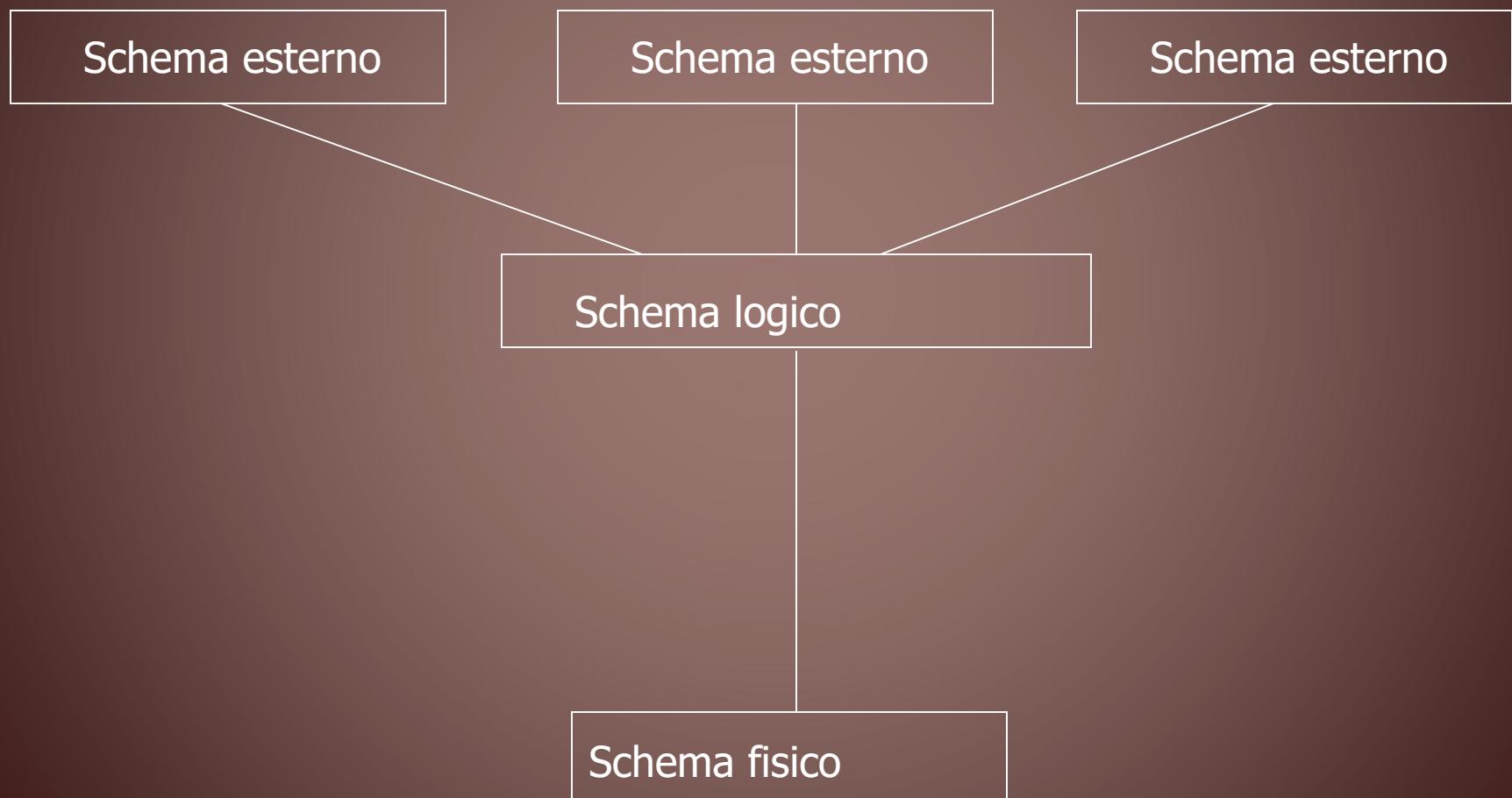


Architetture dati e funzioni

Architetture dei DDBMS

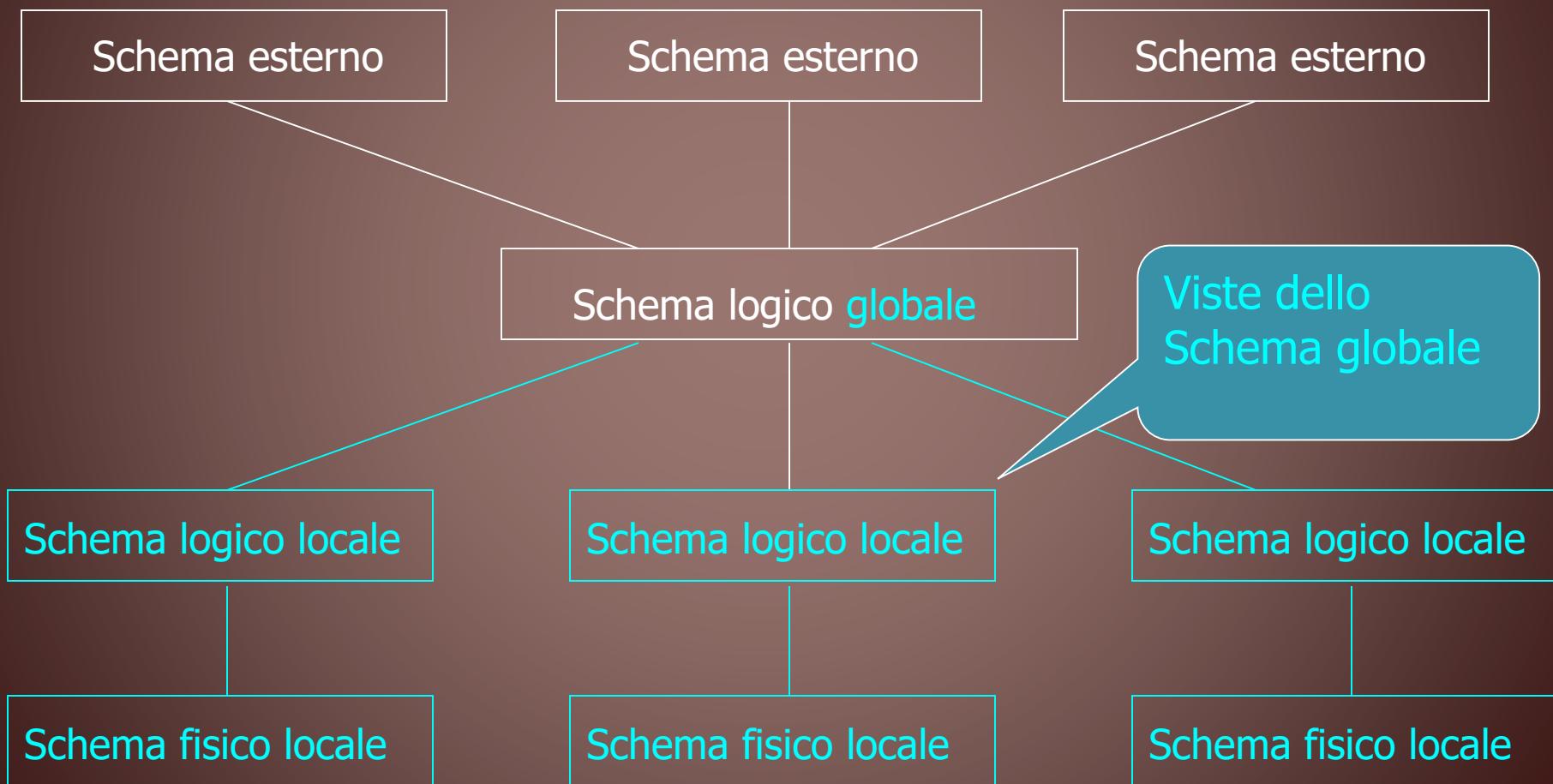
- In particolare vediamo:
 - Per i DDBMS
 - Architettura dati
 - Architettura funzionale

Architettura dati di riferimento per un DBMS centralizzato

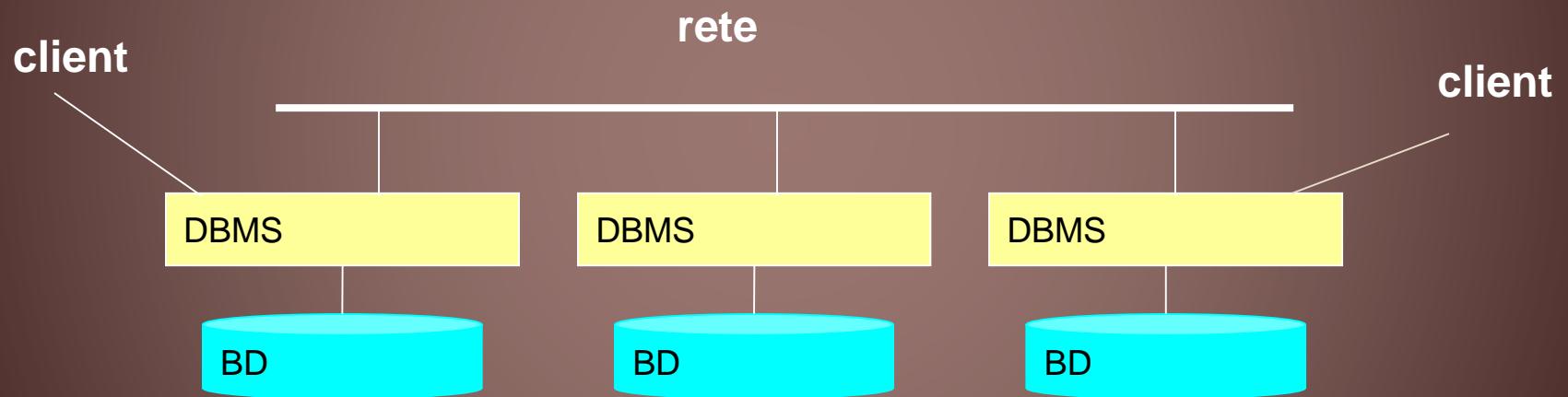


Distributed DBMS

Architettura dati di riferimento per un DDBMS



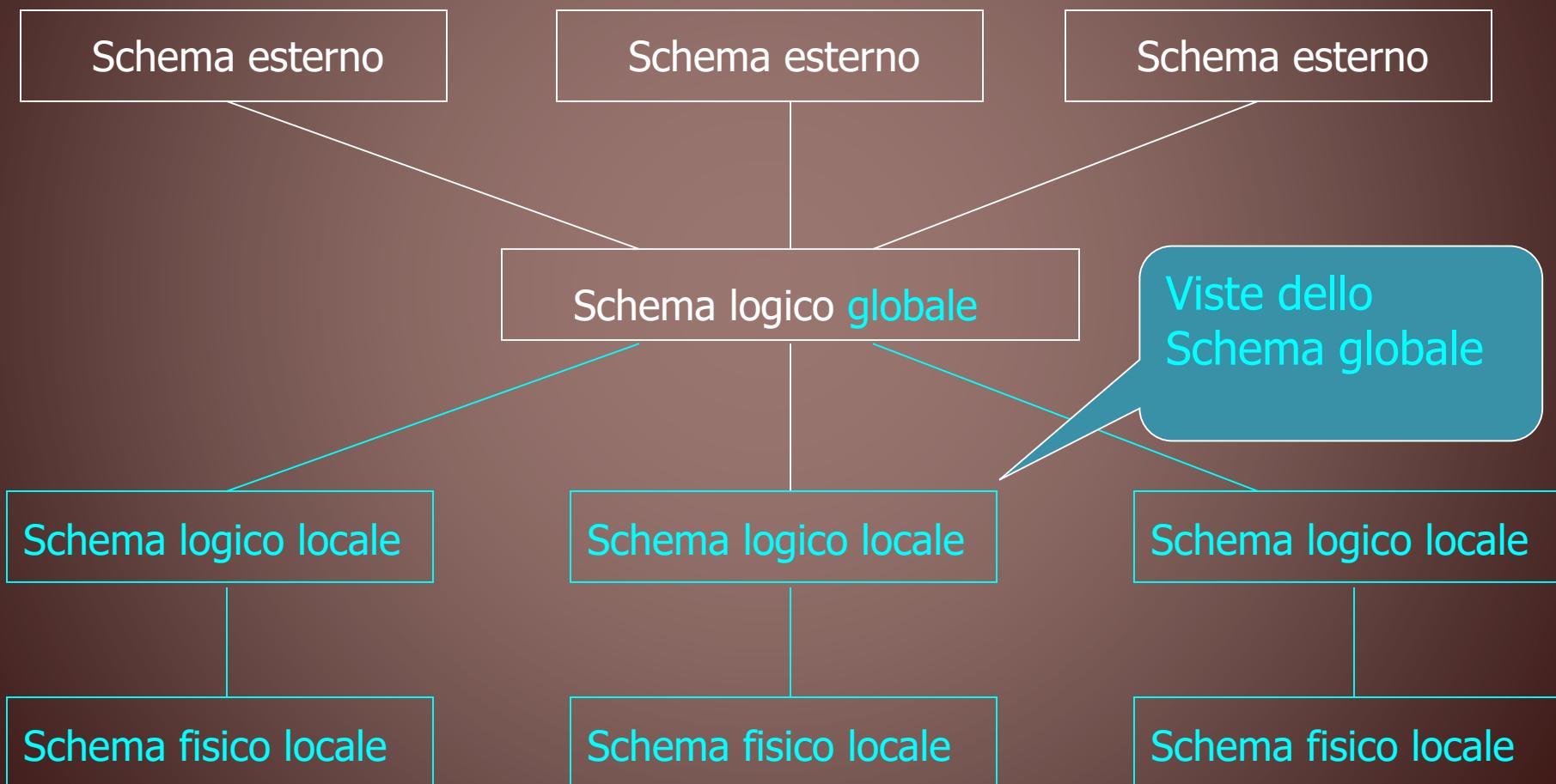
Architettura funzionale di un DDBMS



Ruolo dei DBMS ai diversi nodi

- Per ciascuna funzione (query processing, transaction manager, ecc) vi puo' essere una gestione:
- Centralizzata/gerarchica vs distribuita
- Con assegnazione statica vs dinamica dei ruoli
- Vedremo piu' avanti (DDBMS) esempi concreti di strategie per le diverse funzionalita'

Nei DDBMS gli schemi locali sono viste dello schema globale e vengono progettate, perciò lo schema globale viene progettato prima degli schemi locali





Il processo di progettazione

Progettazione DDBMS

- La natura fortemente integrata dei DDBMS porta a adottare per essi un approccio **top-down** alla progettazione, che rispetto alla progettazione di applicazioni DBMS ...



Progettazione DDBMS

- La natura fortemente integrata dei DDBMS porta ad adottare per essi un approccio top-down alla progettazione, che rispetto alla progettazione di applicazioni DBMS, **introduce una nuova fase**



Progettazione DDBMS

- Inoltre, progettazione logica e fisica sono ora effettuate **sui nodi locali**



Portabilita' e interoperabilita' nei DBMS DEA

- Portabilita' : capacita' di eseguire le stesse applicazioni DB su ambienti runtime diversi
 - Compile time
 - Facilitata dall'aderenza a standards (e.g.: SQL-2, SQL-3)

Portabilita' e interoperabilita' nei DBMS DEA

- **Interoperabilita'**: capacita' di eseguire applicazioni che coinvolgono contemporaneamente sistemi diversi ed eterogenei
 - Run-time
 - Facilitata dalla standardizzazione dei protocolli di accesso ai dati:
 - **Database Connectivity** (ODBC)
 - **X-Open Distributed Transaction Processing** (DTP)
 - Normalmente limitata al livello di accesso

Paradigmi per la distribuzione dati

- Abbiamo visto → **Basi di Dati DEA**, casi DDBMS e Multidatabase: diversi server DBMS utilizzati dallo stesso insieme di applicazioni
- Piu' avanti →
- **Basi di Dati Parallelle**: incremento delle prestazione mediante parallelismo sia di storage devices che di processore
 - **Basi di dati GRID**
- **Basi di Dati Replicate**: replicazione della stessa informazione su diversi server per motivi di performance
- **Data warehouses**: DBMS centralizzati, risultato della integrazione di fonti eterogenee, dedicati specificamente alla gestione di dati per il supporto alle decisioni

Piano di lavoro

	DDBMS	BD Parallele	BD Replicate
Caratteristiche generali	✗	✗	✗
Architetture dati funzioni	✗		
Processo di progettazione	✗		
Proprieta'	✗		
Query processing	✗		
Controllo concorrenza	✗		
Recovery management	✗		