

GESTIONE DI DATI DISTRIBUITI

Prof. Fabio A. Schreiber

Dipartimento di Elettronica e Informazione
Politecnico di Milano



LE TELECOMUNICAZIONI AZIENDALI

1° FASE (< ANNI '80)

- TELEFONO (POTS), TELEX, TRASMISSIONE DATI
INDIPENDENTI SU RETI SEPARATE

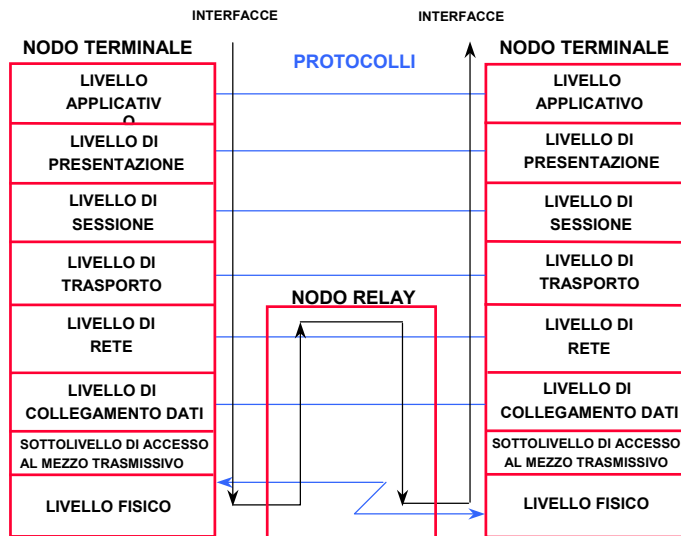
2° FASE (ANNI '80 - '95)

- **PROGETTO E REALIZZAZIONE** DI GRANDI RETI DI TELECOMUNICAZIONE DIGITALI
- SVILUPPO DI **RETI LOCALI** DI PC E WORKSTATION

3° FASE (> 1995)

- **INTEGRAZIONE E GESTIONE** DI GRANDI RETI ETEROGENEE E DI RETI LOCALI

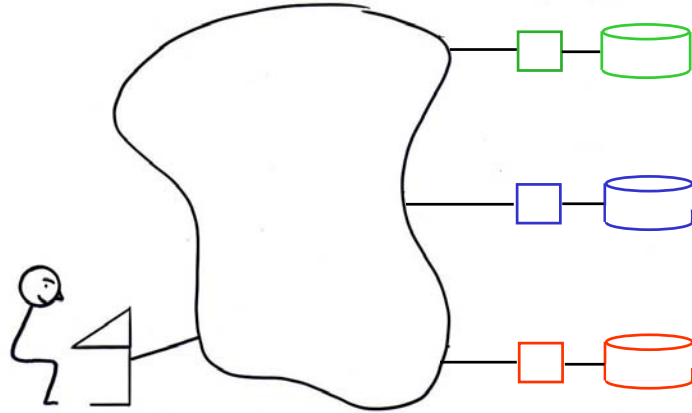
IL MODELLO DI RIFERIMENTO ISO-OSI



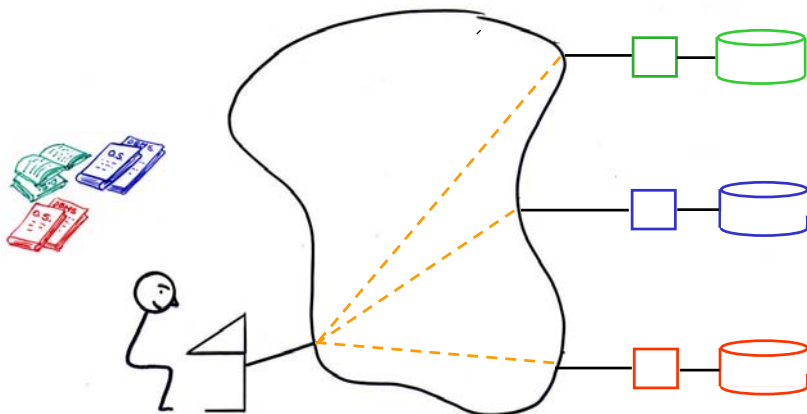
APPLICAZIONI DEI SISTEMI DISTRIBUITI

- **ACCESSO A RISORSE REMOTE**
possibilità di collegarsi mediante **lo stesso terminale** a diversi elaboratori e centri di calcolo (p.e. TELNET, FTP, ...)
- **ELABORAZIONE DISTRIBUITA**
possibilità di realizzare sistemi complessi nei quali **il processo applicativo utilizza più elaboratori e/o più insiemi di dati**, tra loro collegati con sistemi di telecomunicazione (p.e. sistemi informativi distribuiti, HPC, ...)
- **APPLICAZIONI TELEMATICHE**
 - posta elettronica
 - teleconferenza
 - accesso a banche di dati

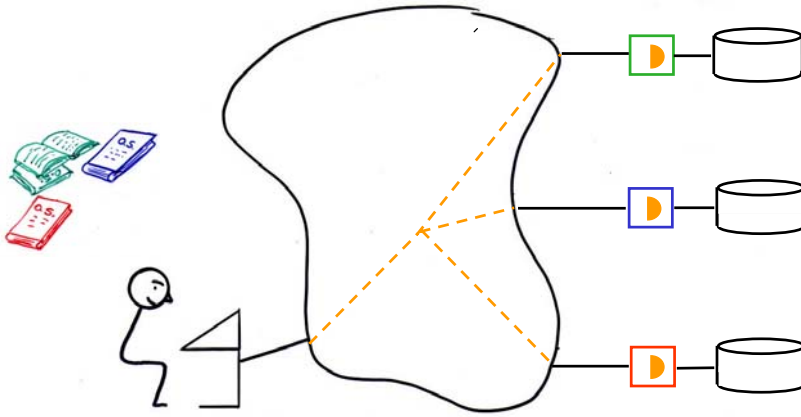
GESTIONE DEI DATI IN RETE



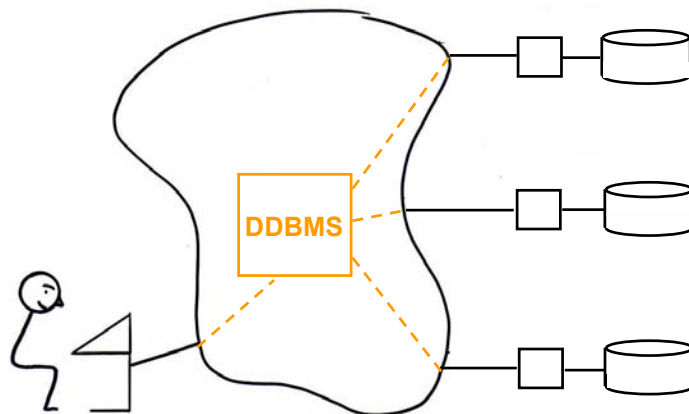
GESTIONE DEI DATI IN RETE: BASI DI DATI INDIPENDENTI



GESTIONE DEI DATI IN RETE: LINGUAGGIO DI COMANDI STANDARD



GESTIONE DEI DATI IN RETE: BASE DI DATI DISTRIBUITA

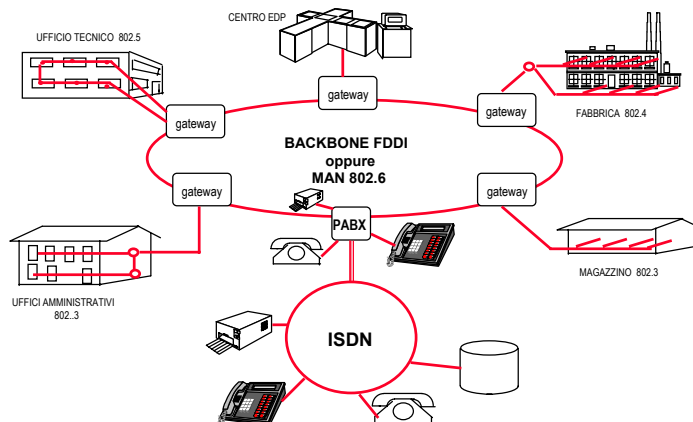


I NUOVI SISTEMI INFORMATIVI

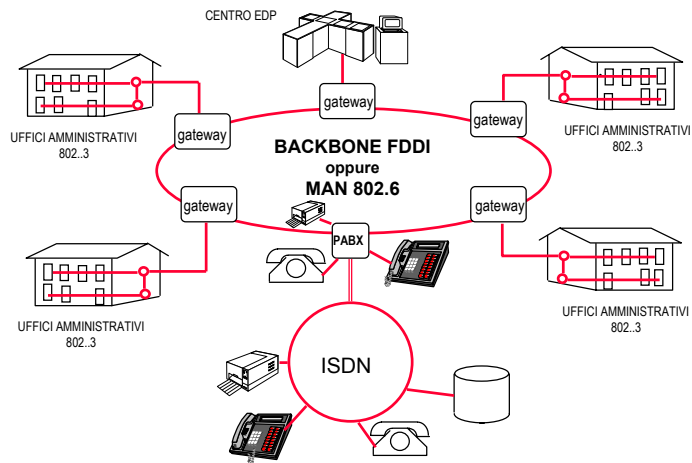
LE RETI DI TELECOMUNICAZIONE DIVENTANO
UNA COMPONENTE ESSENZIALE PER IL **BUON
FUNZIONAMENTO** ECONOMICO E OPERATIVO
DELL'AZIENDA

LA DISPONIBILITA' DI SISTEMI DI TELECOMU-
NICAZIONE EFFICACI CONSENTE LO **SVILUPPO**
DI NUOVI TIPI DI BUSINESS

SISTEMA CON DISTRIBUZIONE VERTICALE DELLE FUNZIONI



SISTEMA CON DISTRIBUZIONE ORIZZONTALE DELLE FUNZIONI



© Gabio A. Schreiber

database distribuiti 10

GESTIONE DEI DATI

1° FASE (< ANNI 70)

- FILE SPARSI

2° FASE (ANNI '70 - '90)

- GRANDI BASI DI DATI CENTRALIZZATE

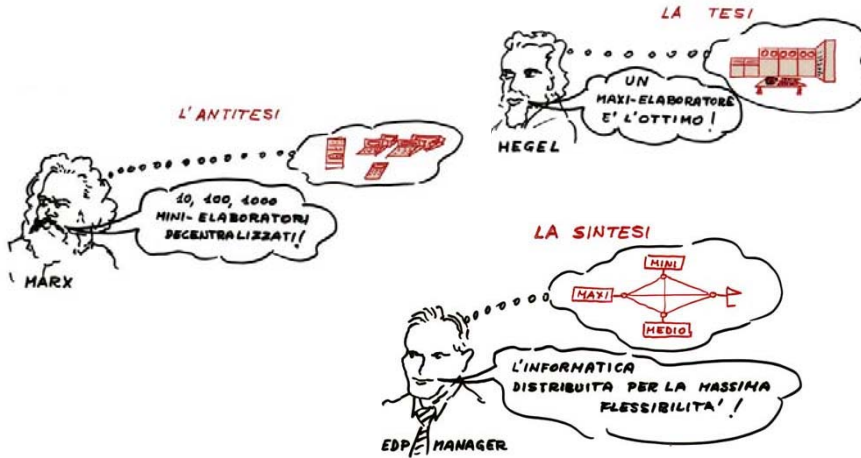
3° FASE (> 1990)

- GESTIONE DISTRIBUITA DEI DATI

© Gabio A. Schreiber

database distribuiti 11

ARCHITETTURA DEI SISTEMI UN PROCESSO DIALETTICO



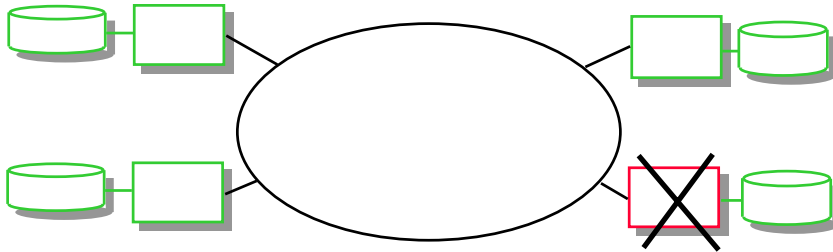
GESTIONE DISTRIBUITA DEI DATI: SCOPI FUNZIONALI

- AFFIDABILITA'
- CONDIVISIONE DEL CARICO DI LAVORO
- CONDIVISIONE DI RISORSE
- QUALITA' DI SERVIZIO PER L'UTENTE

SCOPI FUNZIONALI: AFFIDABILITA'

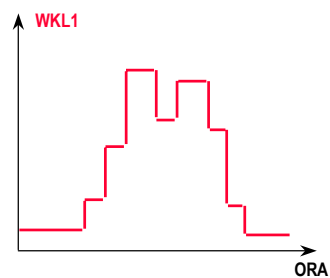
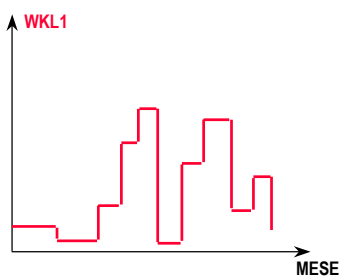
• LA PRESENZA DI RISORSE HW/SW RIDONDANTI CONSENTE DI UTILIZZARLE PER OTTENERE UN' **ELEVATA DISPONIBILITA'** DEL SISTEMA

- SISTEMI **TOLLERANTI AI GUASTI** (FAULT TOLERANT)
- SISTEMI A **DEGRADAZIONE GRADUALE** (SOFT DEGRADATION)



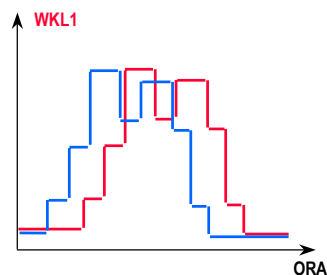
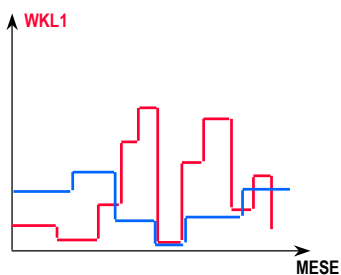
SCOPI FUNZIONALI: CONDIVISIONE DEL CARICO

CONSENTE UN DIMENSIONAMENTO EQUILIBRATO DELLE RISORSE



SCOPI FUNZIONALI: CONDIVISIONE DI CARICO

CONSENTE UN DIMENSIONAMENTO EQUILIBRATO DELLE RISORSE



GESTIONE DISTRIBUITA DEI DATI: SCOPI FUNZIONALI

- **CONDIVISIONE DI RISORSE**
POSSIBILITA' DI CONDIVIDERE RISORSE SPECIALIZZATE O UNICHE PRESENTI IN QUALCHE SITO DEL SISTEMA
- **MIGLIORAMENTO DELLA QUALITA' DEL SERVIZIO**
 - POSSIBILITA' DI ELABORAZIONE LOCALE
 - RIDUZIONE DEI TEMPI DI RISPOSTA
 - INTERFACCIA UTENTE PIU' COMODA

GESTIONE DISTRIBUITA DEI DATI: FATTORI ECONOMICI E ORGANIZZATIVI

- **A FAVORE**

- MAGGIORE CONTATTO TRA UTENTE E SISTEMA GRAZIE ALL'EFFICACIA DEI SISTEMI LOCALI
- DISTRIBUZIONE DEL “POTERE” (ORGANIZZATIVO, PSICOLOGICO, POLITICO, ...) ASSOCIATO AL POSSESSO DELL'INFORMAZIONE
- INTEGRAZIONE DELLE ATTIVITA' ORGANIZZATIVE IN IMPRESE TERRITORIALMENTE DISTRIBUITE
- COSTI/BENEFICI ???

- **CONTRO**

- LA NECESSITA' DI COORDINAMENTO E COLLABORAZIONE RICHIESTA A TUTTI I LIVELLI RICHIEDA UNA “ATTITUDINE CULTURALE” CHE NON SI TROVA FREQUENTEMENTE
- COSTI/BENEFICI ???

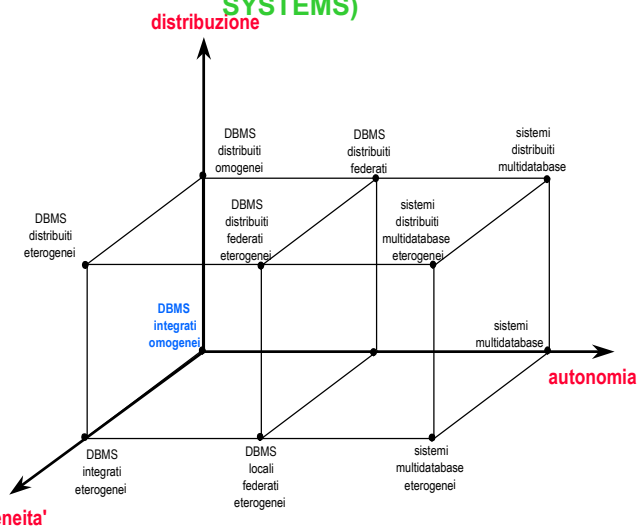
SISTEMI DISTRIBUITI

SONO SISTEMI NEI QUALI IL TEMPO DI TRASMISSIONE DEI MESSAGGI **NON E' TRASCURABILE** RISPETTO AL TEMPO CHE INTERCORRE TRA GLI EVENTI DI UN SINGOLO PROCESSO

GESTIONE DISTRIBUITA DEI DATI: ARCHITETTURE

ARCHITETTURA Hw/Sw	GESTIONE DEI DATI
sistemi debolmente connessi a banda stretta reti di calcolatori geograficamente distribuite	sistemi per la gestione di basi di dati distribuite
sistemi debolmente connessi a banda larga reti locali, sistemi funzionalmente distribuiti	back-end processor
sistemi paralleli fortemente connessi a banda larga sistemi multiprocessori, memorie associative, ...	database machine

LO SPAZIO DEI DDSS (DISTRIBUTED DATA SHARING SYSTEMS)



TASSONOMIA DEI DDSS

TIPO DI SISTEMA	IL SISTEMA GLOBALE ACCEDE A ...	I NODI LOCALI SONO <small>TIPICAMENTE ...</small>	FUNZIONI DI DB GLOBALE	MODO DI PORRE L'INFORMAZ. GLOBALE
basi di dati distribuite	funzioni interne del DBMS	basi di dati omogenee	si	spazio dei nomi e schema globali
multidatabase con schema globale	interfaccia utente del DBMS	basi di dati eterogenee	si	schema globale
basi di dati federate	interfaccia utente del DBMS	basi di dati eterogenee	si	schemi globali parziali
multidatabase con linguaggio di sistema	interfaccia utente del DBMS	basi di dati eterogenee	si	funzioni di accesso del linguaggio
multidatabase omogenei con linguaggio di sistema	interfaccia utente del DBMS e alcune funzioni interne	basi di dati omogenee	si	funzioni di accesso del linguaggio
sistemi interoperabili	applicazioni al di sopra del DBMS	qualsiasi sorgente di dati che soddisfi il protocollo di interfaccia	no	scambio di dati

GESTIONE DISTRIBUITA DEI DATI

- **ARCHITETTURA CLIENT-SERVER**
 - **MOLTI** CLIENTI SI RIVOLGONO AD **UN UNICO** SERVER
 - USATA PRINCIPALMENTE PER **OLTP** SU RETI LOCALI

- **BASI DI DATI DISTRIBUITE**
 - **MOLTI** CLIENTI SI RIVOLGONO A **MOLTI** SERVERNTI
 - USATE PER **OLTP** SU RETI LOCALI E GEOGRAFICHE

- **DATA WAREHOUSE**
 - COLLEZIONE DI DATI PROVENIENTI DA PIU' SISTEMI SORGENTE
 - USATE PER **DSS** SU RETI LOCALI E GEOGRAFICHE

DDSS IN ARCHITETTURA CLIENT-SERVER

• IL CLIENT

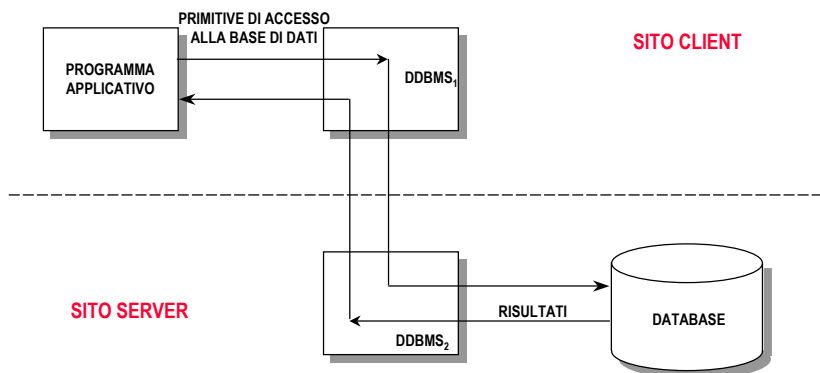
- E' GESTITO DAL PROGRAMMATORE APPLICATIVO
- E' ADATTO ALL'INTERAZIONE CON L'UTENTE FINALE AL QUALE PRESENTA UN'INTERFACCIA AMICHEVOLE
- UTILIZZA SQL STATICO O DINAMICO

• IL SERVER

- E' GESTITO DALL' AMMINISTRATORE DELLA BD
- DIMENSIONATO IN FUNZIONE DEL CARICO E DEI SERVIZI CHE DEVE OFFRIRE
- GESTISCE LE PROCEDURE DI OTTIMIZZAZIONE

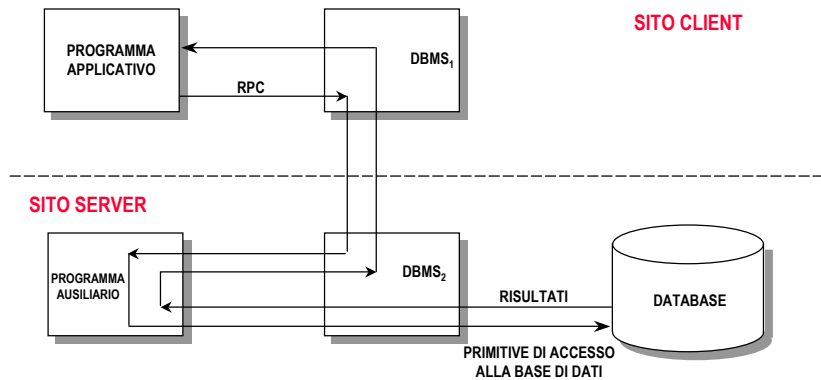
DDSS IN ARCHITETTURA CLIENT-SERVER

USO DI PRIMITIVE DEL DDBMS



DDSS IN ARCHITETTURA CLIENT-SERVER

USO DI PROGRAMMI AUSILIARI E RPC



DDSS IN ARCHITETTURA CLIENT-SERVER

• USO DI PRIMITIVE DEL DDBMS

- IL COMPONENTE LOCALE DEL DDBMS INDIRIZZA LA RICHIESTA AL SERVER CHE ESEGUE L'ACCESSO E RESTITUISCE I RISULTATI
- MAGGIOR **TRASPARENZA** ALLA DISTRIBUZIONE GRAZIE A NOMI DI FILE GLOBALI
- **BASSA EFFICIENZA** PERCHE' VIAGGIA UNA TUPLA ALLA VOLTA

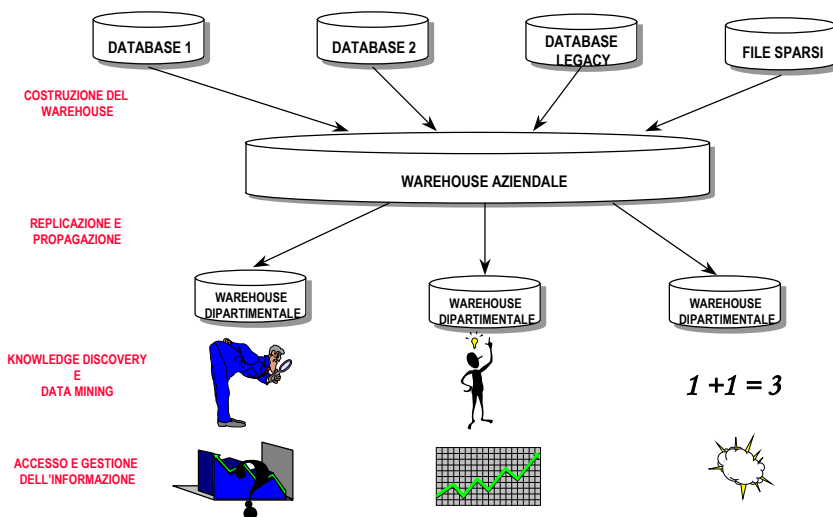
• USO DI PROGRAMMI AUSILIARI E RPC

- L'APPLICAZIONE CHIEDE L'ESECUZIONE DEL PROGRAMMA AUSILIARIO SUL SERVER. QUESTO ESEGUE L'OPERAZIONE E RESTITUISCE I RISULTATI
- IL PROGRAMMA AUSILIARIO E' IN GRADO DI ASSEMBLARE LE SINGOLE TUPLE IN INSIEMI RISULTATO COSI' **AUMENTANDO L'EFFICIENZA** DELLE TRASMISSIONI

DATA WAREHOUSE (DW)

- TECNICA PER ASSEMBLARE E GESTIRE CORRETTAMENTE DATI PROVENIENTI DA SORGENTI DIVERSE AL FINE DI OTTENERE UNA VISIONE DETTAGLIATA DI UN SISTEMA ECONOMICO
- E' UNA RACCOLTA DI DATI
 - INTEGRATA
 - PERMANENTE
 - VARIABILE NEL TEMPO
 - ORIENTATA AD UN PRECISO ARGOMENTOA SUPPORTO DI DECISIONI MANAGERIALI
- E' L'ELEMENTO DI SEPARAZIONE TRA I CARICHI DI LAVORO OLTP E QUELLI DSS

STRUTTURA DI UN DW



PRINCIPALI PROBLEMATICHE DI UN DW

- MANUTENZIONE DELLE **VIEW** E DEI **METADATI**
- GESTIONE DELLA **REPLICAZIONE**
- GESTIONE DELLA **CONSISTENZA**
- REALIZZAZIONE DELLE **APPLICAZIONI**

UNA BASE DI DATI DISTRIBUITA

E' UN INSIEME DI FILE, MEMORIZZATI IN NODI DIVERSI DI UN SISTEMA DISTRIBUITO, CHE SONO **LOGICAMENTE CORRELATI** O DA RELAZIONI FUNZIONALI O PERCHE' COPIE MULTIPLE DI UNO STESSO FILE, IN MODO DA COSTITUIRE **UN'UNICA COLLEZIONE DI DATI**

UNA BASE DI DATI DISTRIBUITA ...

- **E' UNA BASE DI DATI**
 - DEVE ESISTERE UN **MODO INTEGRATO PER ACCEDERE AI DATI**
 - DEVE ESSERE PROTETTA CONTRO INCONSISTENZE E GUASTI IN MODO DA **GARANTIRE L'INTEGRITA' DEI DATI**
- **E' DISTRIBUITA**
 - LA **DISTRIBUZIONE FISICA DEI DATI** DEVE ESSERE **TRASPARENTE** ALL'UTENTE FINALE

ALCUNI PROBLEMI DI PROGETTO

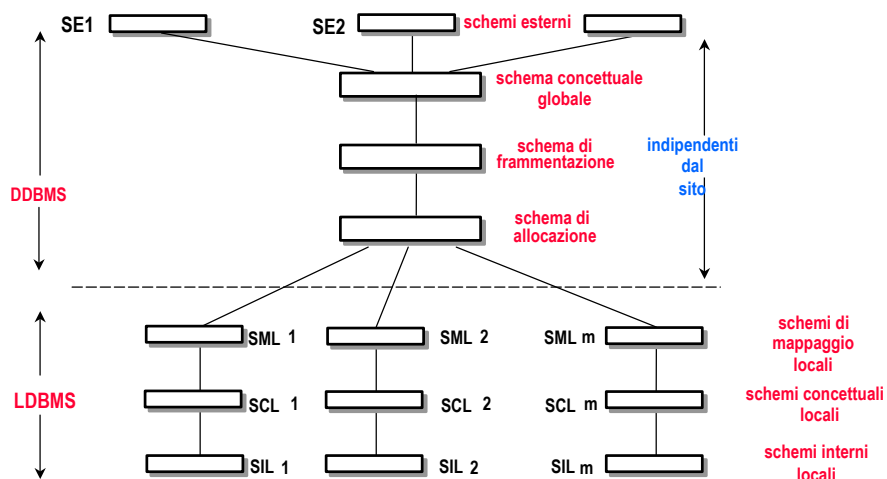
- **ARCHITETTURA** GENERALE DEL SISTEMA
 - PROGETTO EX-NOVO
 - RISTRUTTURAZIONE DI SISTEMI ESISTENTI
 - ETEROGENEITA' DEI SISTEMI E DEI DATI
- **FRAMMENTAZIONE** DELLE RELAZIONI LOGICHE
- **REPLICAZIONE E ALLOCAZIONE**
 - QUANTE COPIE E DOVE
- **ACCESSO ED ELABORAZIONE** DELLE RELAZIONI
- **GARANZIA DI INTEGRITA' E PRIVATEZZA**
- **GARANZIA DI AFFIDABILITA'**

INDIPENDENZA DEI DATI

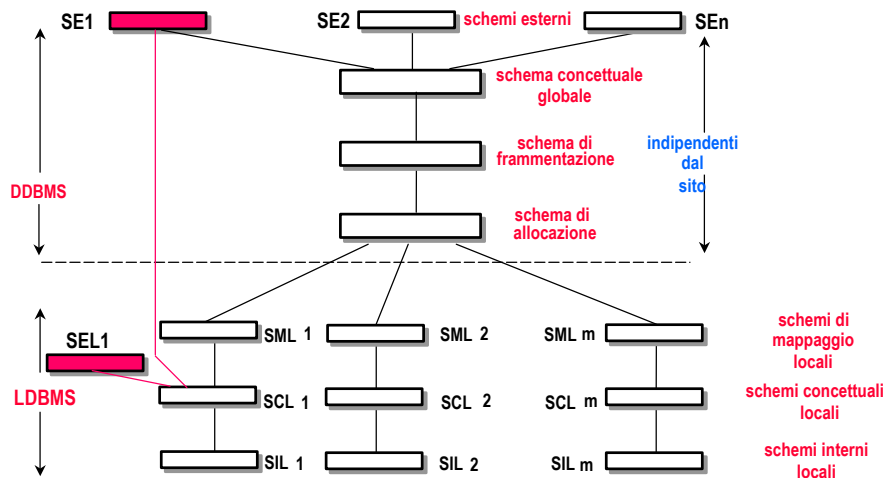
IL CONCETTO DI INDIPENDENZA DEI DATI VIENE ESTESO PER COMPRENDERE I SEGUENTI CASI

- **LOGICA**
NECESSITA' DELL'AMMINISTRATORE DELLA BD DI ADATTARE LO SCHEMA GLOBALE AI REQUISITI DI UN INSIEME DI UTENTI MOLTO GRANDE, ETEROGENEO E DINAMICO
- **FISICA**
IMMUNITA' AI CAMBIAMENTI - ANCHE DINAMICI - DI CONFIGURAZIONE DELLA RETE (PRESENZA O ASSENZA DI ALCUNI SITI)

DDBMS INTEGRATO



DDSS FEDERATO



© Gabio A. Schreiber

database distribuiti 36

PROGETTO DI SCHEMA GLOBALE

- ANALOGO ALL'INTEGRAZIONE DI VIEW
- CONFLITTI STRUTTURALI (schemi differenti)
- CONFLITTI SEMANTICI (anche con schemi uguali)
 - **OMONIMIA** (STESSO NOME, MA SIGNIFICATO DIVERSO)
 - **AMBIGUITA'** (NOME DIVERSO, MA STESSO SIGNIFICATO)
 - **FORMATO** (NUMERICO ↔ ALFANUMERICO, ...)
 - **AGGREGAZIONI** (PART-OF, ...)
 - **SEMANTICA PROPRIA DEL DATO** (DIFFERENZA DI UNITA' DI MISURA, DI VALUTAZIONE, DI GRANULARITA', ...)
- AGGIORNAMENTO SPESSO RISTRETTO ALL'ATTIVITA' **LOCALE**

© Gabio A. Schreiber

database distribuiti 37

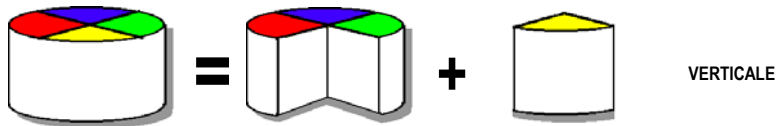
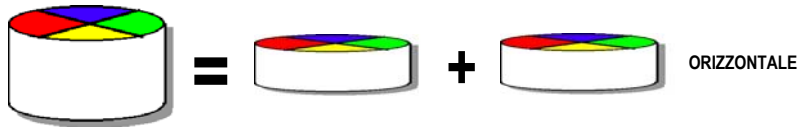
RISOLUZIONE DEI CONFLITTI SEMANTICI

- RISTRUTTURAZIONE DELLE BASI DI DATI LOCALI (POCO PROPONIBILE)
- AGGIUNTA **ESPLICITA** NEGLI SCHEMI DI INFORMAZIONI SULLA SEMANTICA DEI SINGOLI DATI IN MODO DA PERMETTERE ALLE **PROCEDURE APPLICATIVE** DI FARVI FRONTE

FRAMMENTAZIONE DELLE RELAZIONE LOGICHE

- **ORIZZONTALE**
 - TUTTI I FRAMMENTI HANNO LO STESSO **SCHEMA**
 - L'ASSEGNAZIONE DELLE TUPLE AI FRAMMENTI AVVIENE IN BASE AD UN **PREDICATO DI SELEZIONE** CHE CORRISPONDE AD UN **CRITERIO DI DISTRIBUZIONE**
 - CONSENTE DI REALIZZARE SISTEMI NEI QUALI **IL VALORE DI UN DATO DIPENDE DALLA SUA LOCAZIONE**
- **VERTICALE**
 - OGNI FRAMMENTO HA UNO SCHEMA OTTENUTO PER **PROIEZIONE** DALLO SCHEMA DELLA RELAZIONE GLOBALE
 - SCHEMI CON **INTERSEZIONE NON NULLA**
 - SCHEMI **DISGIUNTI**

FRAMMENTAZIONE DELLE RELAZIONE LOGICHE



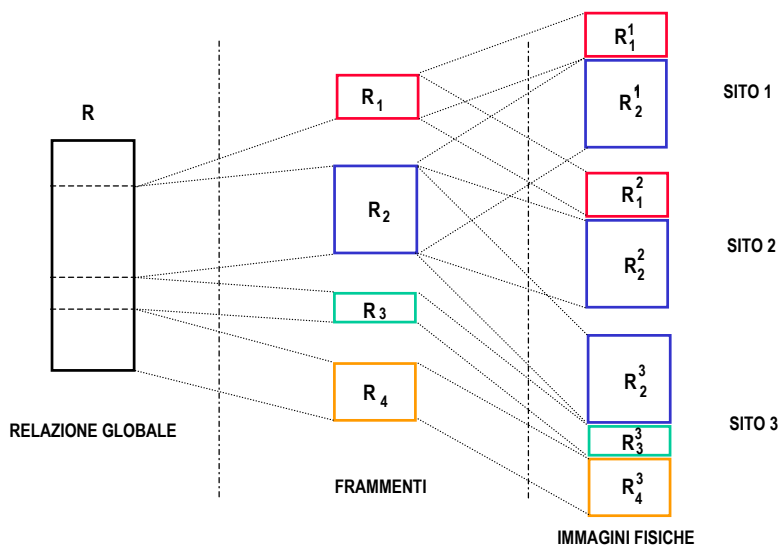
REPLICAZIONE DEI DATI

- **PERMANENZA**
 - LA COPIA DI UN ELEMENTO DI DATI (RELAZIONE O FRAMMENTO) E' **PERMANENTE** SE EVOLVE NEL TEMPO SOTTO LA GESTIONE DEL DDBMS
 - E' INVECE **TEMPORANEA** SE ESSA VIENE CREATA SOLAMENTE PER QUALCHE OPERAZIONE SPECIFICA (IN UN'AREA DI LAVORO) E QUINDI VIENE CANCELLATA O RIGENERATA, SU RICHIESTA, A PARTIRE DALL'ORIGINALE

REPLICAZIONE DEI DATI

- **CONSISTENZA**
 - **CONSISTENZA FORTE (DIPENDENZA COMPLETA)**
IN OGNI ISTANTE OGNI COPIA DI CIASCUN ELEMENTO DI DATI DEVE AVERE IDENTICO VALORE
 - **CONSISTENZA DEBOLE (DIPENDENZA RIDOTTA)**
GLI AGGIORNAMENTI APPORTATI AD UNA COPIA VENGONO PROPAGATI ALLE ALTRE IN MODO DIFFERITO
 - **INDIPENDENZA**
GLI AGGIORNAMENTI SULLE DIVERSE COPIE SONO SCORRELATI (SPESSO USATA CON COPIE TEMPORANEE)

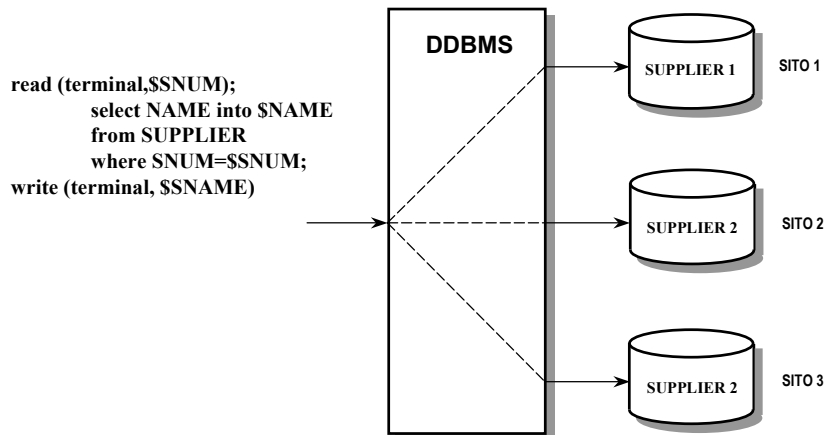
FRAMMENTAZIONE E REPLICAZIONE



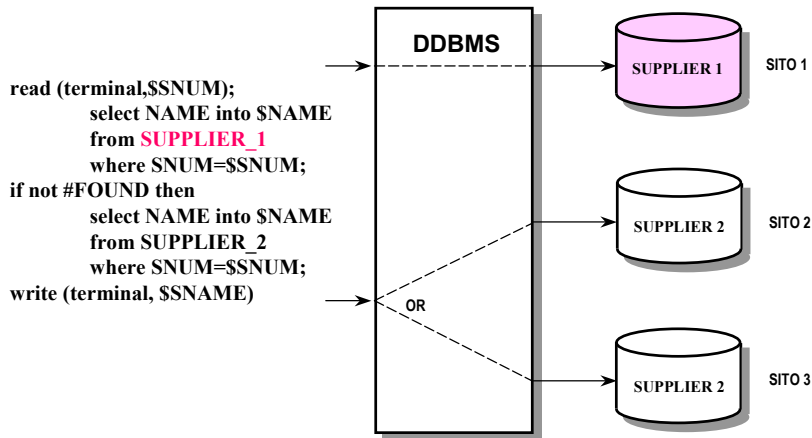
LIVELLI DI TRASPARENZA

- **ALLA FRAMMENTAZIONE**
 - I PROGRAMMI APPLICATIVI FANNO RIFERIMENTO ALLE RELAZIONI GLOBALI IGNORANDO LA FRAMMENTAZIONE
- **ALLA ALLOCAZIONE**
 - I PROGRAMMI APPLICATIVI SONO INDIPENDENTI DALLA REPLICAZIONE E DALLA ALLOCAZIONE FISICA DEI DATI, MA RISENTONO DI VARIAZIONI DELLO SCHEMA DI FRAMMENTAZIONE
- **AL MAPPING LOCALE**
 - L'APPLICAZIONE USA I NOMI GLOBALI DEGLI OGGETTI (FRAMMENTI DI DATI O PRIMITIVE DI ACCESSO), MA DEVE SPECIFICARE IL SITO DI RESIDENZA
- **NESSUNA TRASPARENZA**
 - IL PROGRAMMATORE APPLICATIVO DEVE CODIFICARE I SINGOLI MODULI DI ACCESSO PER I VARI DBMS, CHE SI LIMITANO AD ATTIVARE I MODULI REMOTI

LIVELLI DI TRASPARENZA: INTERROGAZIONI trasparenza alla frammentazione



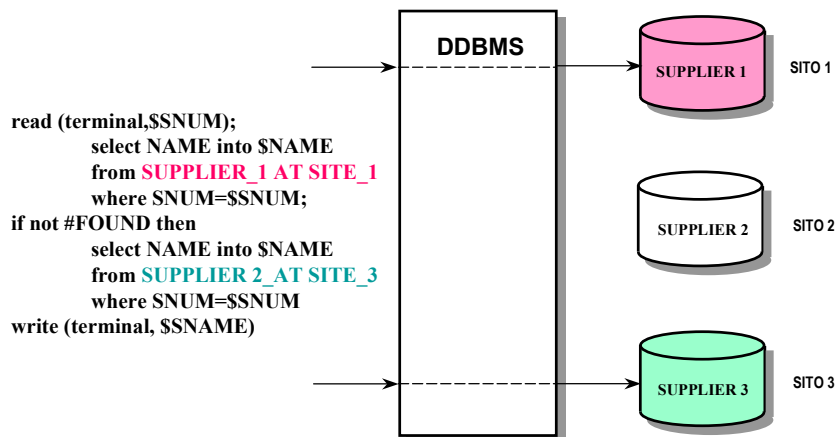
LIVELLI DI TRASPARENZA: INTERROGAZIONI trasparenza all'allocazione



©abio A. Schreiber

database distribuiti 46

LIVELLI DI TRASPARENZA: INTERROGAZIONI trasparenza al mapping locale

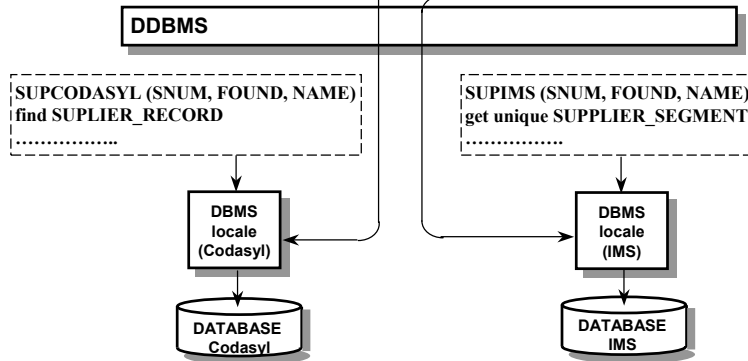


©abio A. Schreiber

database distribuiti 47

LIVELLI DI TRASPARENZA: INTERROGAZIONI nessuna trasparenza

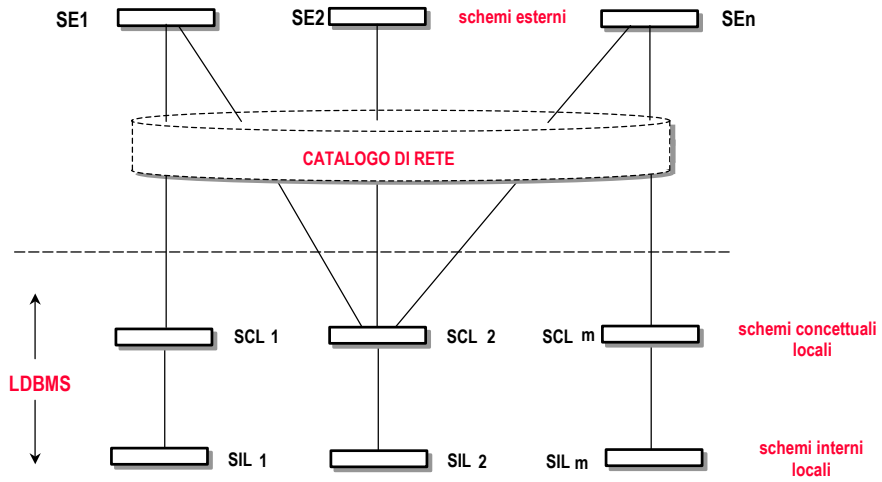
```
SUPPINQRY:
read (terminal, $SNUM);
  execute $SUPIMS ($SNUM, $FOUND, $NAME) at site 1;
if not $FOUND then
  execute $SUPCODASYL ($SNUM, $FOUND, $NAME) at site 3;
write (terminal, $NAME)
```



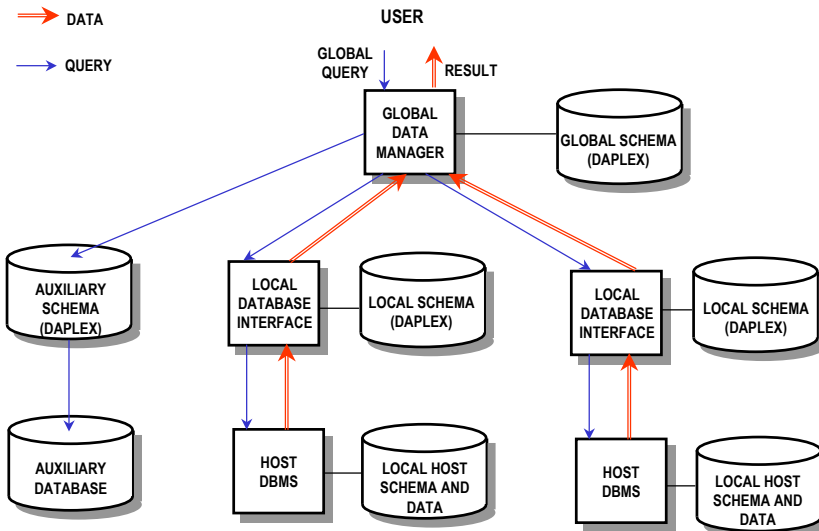
LIVELLI DI TRASPARENZA: AGGIORNAMENTO

- PER LA TRASPARENZA ALLA **FRAMMENTAZIONE**
 - SE SI AGGIORNA UN ATTRIBUTO CHE COMPARE IN UN PREDICATO DI FRAMMENTAZIONE ORIZZONTALE, OCCORRE **SPOSTARE LA TUPLA DA UN FRAMMENTO AD UN ALTRO**
- PER LA TRASPARENZA ALL'**ALLOCAZIONE** E ALLA **REPLICAZIONE**
 - DEVONO ESSERE **AGGIORNATE TUTTE LE COPIE**

MULTIDATABASE



MULTIBASE



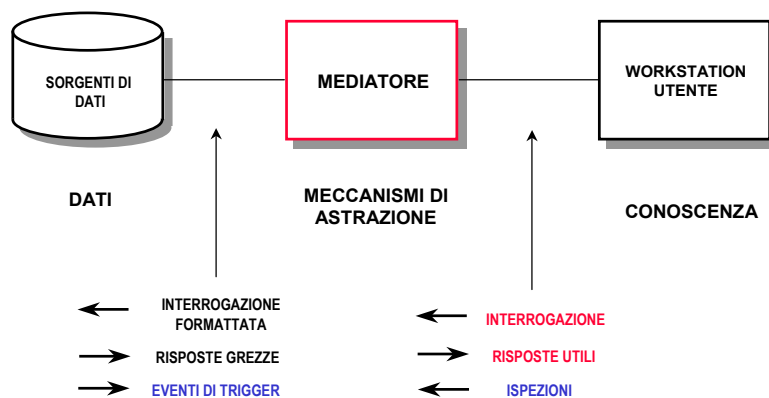
MEDIATORI

UN MEDIATORE E' UN MODULO SOFTWARE (A LIVELLO 7 ISO-OSI) CHE UTILIZZA LA CONOSCENZA IMPLICITA DI ALCUNI INSIEMI O SOTTOINSIEMI DI DATI PER CREARE INFORMAZIONE PER UN LIVELLO SUPERIORE DI APPLICAZIONI (WIEDERHOLD)

LA SUA FUNZIONE PRINCIPALE E' LA **FUSIONE** DI OGGETTI

- **RAGGRUPPARE** LE INFORMAZIONI SULLA STESSA ENTITA' DEL MONDO REALE
- **RIMUOVERE** LE RIDONDANZE TRA LE DIVERSE SORGENTI
- **RISOLVERE** LE INCONSISTENZE TRA LE DIVERSE SORGENTI

MEDIATORI



WRAPPER

CONVERTONO LE INTERROGAZIONI IN UNO O PIU' COMANDI/INTERROGAZIONI COMPRESIBILI DALLA SPECIFICA SORGENTE

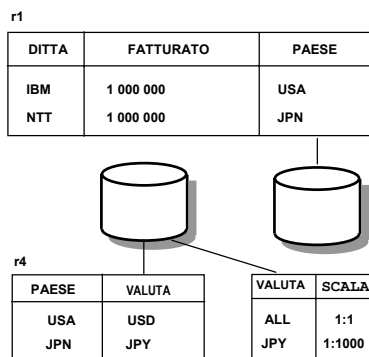
POSSONO **ESTENDERE** LA POSSIBILITA' DI INTERROGAZIONE DI UNA PARTICOLARE SORGENTE

CONVERTONO I RISULTATI DAL FORMATO NATIVO AD UN FORMATO COMPRESIBILE ALL'APPLICAZIONE

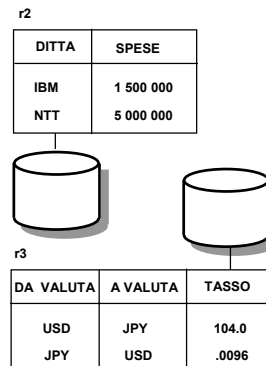
LA LORO SCRITTURA COMPORTA UN NOTEVOLE SFORZO REALIZZATIVO, MA ESISTONO STRUMENTI CHE AGEVOLANO IL LAVORO (TSIMMIS TOOLKIT)

ESEMPIO DI CONFLITTO SEMANTICO

CONTESTO C1:
• VALORI MONETARI IN VALUTA DI ORIGINE
• VALORI MONETARI IN SCALA 1:1 TRANNE LO YEN CHE E' IN SCALA 1:1000

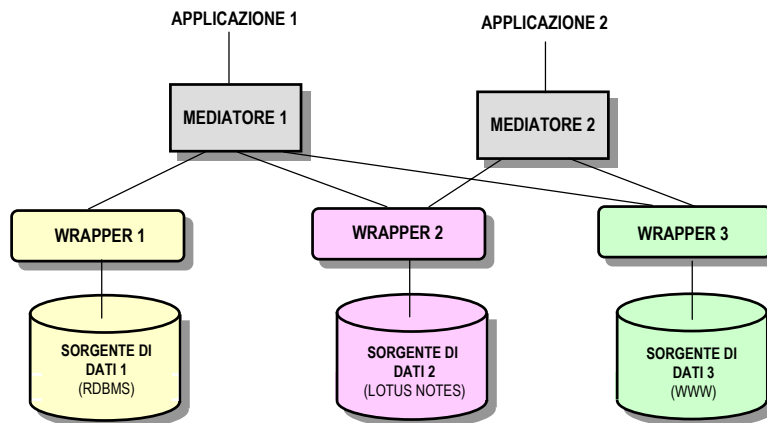


CONTESTO C2:
• VALORI MONETARI IN USD IN SCALA 1:1

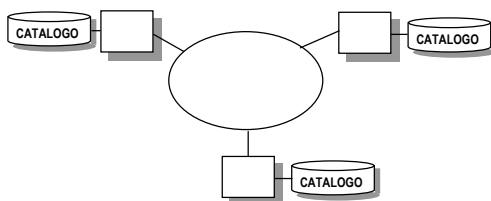


```
select r1.ditta, r1.fatturato
from r1, r2
where r1.ditta = r2.ditta
and r1.fatturato > r2.spese
```

ESEMPIO DI ARCHITETTURA A MEDIATORI (TSIMMIS)

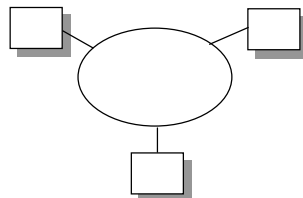


CATALOGO DI RETE



COMPLETAMENTE REPLICATO

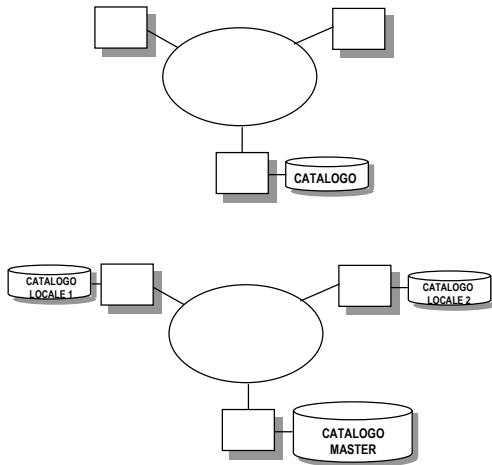
- IMMEDIATEZZA DI ACCESSO
- OCCUPAZIONE DI DISCO
- CONSISTENZA DEI DATI



COMPLETAMENTE ASSENTE

- NECESSITA' DI BROADCASTING
- OVERHEAD DI ACCESSO

CATALOGO DI RETE



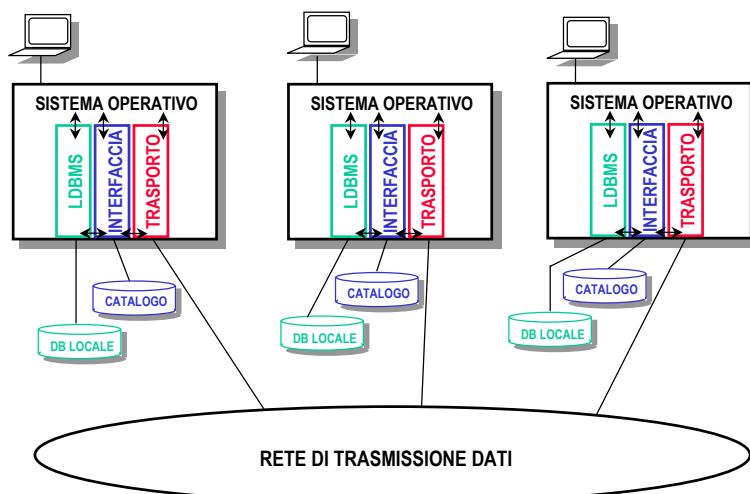
COMPLETAMENTE CENTRALIZZATO

- COLLO DI BOTTIGLIA PER L'ACCESSO
- OCCUPAZIONE DI DISCO
- CONSISTENZA DEI DATI

PARZIALMENTE REPLICATO

- BUON COMPROMESSO TRA OVERHEAD E EFFICIENZA IN ACCESSO

ARCHITETTURA COMPLESSIVA



MODELLO PER SERVIZI DI INFORMAZIONI DISTRIBUITE ED ETEROGENEE

