Tecniche di protezione delle basi di dati

Lezione 24 di Sicurezza dei sistemi informatici 1

Docente: Giuseppe Scollo

Università di Catania, sede di Comiso (RG)
Facoltà di Scienze Matematiche, Fisiche e Naturali
Corso di Studi in Informatica applicata, AA 2008-9

Indice

1. Tecniche di protezione delle basi di dati
2. aggiornamento in due fasi
3. controllo della concorrenza
4. vincoli d'integrità
5. protezione di dati sensibili
6. il problema dell'inferenza
7. inferenza e aggregazione

aggiornamento in due fasi

il registro delle transazioni permette di mantenere l'integrità di una base di dati, dopo il ripristino dal *back-up* più recente, attraverso la riesecuzione delle transazioni **completate** dopo la creazione di quest'ultimo, ma ... come trattare le transazioni **interrotte** ?

l'aggiornamento in due fasi (*two-phase commit*) risolve questo problema:

1. **fase dell'intento** : le operazioni di scrittura non vengono effettuate sulla base di dati, bensì su **variabili-ombra**, corrispondenti ai campi da aggiornare
2. **fase di commit** : i valori delle variabili-ombra vengono copiati nei corrispondenti campi del *record* da aggiornare

un **commit flag** nella base di dati cambia di stato al passaggio dalla prima alla seconda fase, per tornare allo stato iniziale dopo la registrazione del completamento di quest'ultima nel registro delle transazioni

in tal modo, se l'esecuzione della transazione viene interrotta, è possibile rieseguirla, per intero o solo la sua seconda fase a seconda dello stato del **commit flag**, senza pregiudicare integrità e accuratezza dell'informazione presente nella base di dati

caso limite: se l'interruzione si verifica **dopo** la registrazione del completamento della transazione ma **prima** della reinizializzazione del **commit flag**... che fare?

controllo della concorrenza

l'accesso concorrente può essere fonte di violazioni dell'integrità e/o accuratezza dell'informazione presente in una base di dati, per diverse ragioni:

concorrenza di **scrittura** di uno stesso elemento → **perdita di aggiornamenti**

concorrenza di **lettura e scrittura** di un elemento → **incoerenza** dell'informazione letta con quella effettivamente presente nella base di dati

meccanismi di **blocco (lock)** temporaneo dell'accesso risolvono il problema, tuttavia ...

... spesso occorre che il blocco si applichi ad una **sequenza** di operazioni, ovvero ad un'intera **transazione**, per garantirne l'**atomicità**

in questo aspetto, il controllo della concorrenza è simile all'aggiornamento in due fasi, che in effetti garantisce l'atomicità di una transazione anche in presenza di interruzioni della sua esecuzione

soluzione: **two-phase lock**

vincoli d'integrità

i vincoli d'integrità di una base di dati servono a garantirne:

consistenza interna: integrità della struttura logica dei dati

consistenza esterna (accuratezza): veridicità dell'informazione

per la loro finalità, la definizione di adeguati vincoli d'integrità dipende dall'applicazione, ed è dunque una responsabilità del progettista della base di dati, tuttavia ...

due tipici vincoli d'integrità delle basi di dati **relazionali** sono indipendenti dall'applicazione, e possono dunque essere imposti e controllati dal DBMS:

integrità di entità: nessun componente della chiave primaria può essere indefinito (assente)

integrità referenziale: ad ogni valore di chiave esterna deve corrispondere un elemento nella relazione a cui la chiave esterna fa riferimento

vincoli d'integrità specifici per l'applicazione tipicamente sono:

controlli di campo (locali all'elemento)

controlli di consistenza (fra elementi di relazioni diverse)

i DBMS tipicamente forniscono funzionalità per la definizione di tali vincoli, nonché di **procedure di controllo d'integrità** ad essi associate

protezione di dati sensibili

fattori di sensibilità dei dati:

inerente

provenienza da fonte sensibile (informatore, agente sotto copertura, ...)

dichiarati sensibili

parte di un aggregato sensibile (attributo, record, ...)

sensibilità relativa ad altri dati (già divulgati)

fattori di influenza nella decisione di accesso ai dati:

disponibilità dei dati

accettabilità della richiesta di accesso

garanzia di autenticità del richiedente

tipi di divulgazione (parziale) di dati (sensibili)

dati esatti

confini

risultato negativo

esistenza

valore probabile

obiettivi della protezione di dati sensibili: **sicurezza + precisione**

il problema dell'inferenza

inferenza (in questo contesto):

deduzione di dati sensibili da dati non sensibili

attacchi:

query mascherata

attacco indiretto ("statistico")

difesa:

vincoli del tipo: n elementi $> k$ %

controlli per attacchi di inferenza statistica:

sulla **query**: generalmente efficaci solo contro attacchi diretti, non qui sui **dati**:

suppressione (rifiuto della query)

occultamento (risposta imprecisa)

inferenza e aggregazione

il problema dell'inferenza per aggregazione di dati si presenta in due forme, entrambe caratterizzate da un diverso livello di sensibilità L_a del dato aggregato rispetto a quello L_o dei dati oggetto di aggregazione:

$L_a > L_o$: ad es., sommario di dati economici d'impresa aggregato da dati economici dipartimentali

$L_a < L_o$: ad es., valor medio del grado di dipendenza da droghe in una popolazione di individui

la soluzione nel primo caso consta essenzialmente di un'adeguata definizione delle autorizzazioni, combinata con efficaci tecniche di controllo di accesso

il secondo caso è più complesso; i tipi di attacchi accennati in precedenza rientrano tutti in questa categoria, assieme ad altri più sofisticati:

interrogazioni con predicati *tracker*, inferenza per risoluzione di sistemi lineari di equazioni, ...