

Introduzione

1 Modelli dei dati

Oltre a quello relazionale, esistono numerosi altri modelli dei dati, tra cui:

- **object-oriented**,
- **object-relational**,
- **tree-based (semi-structured)**,
- **graph-based**.

1.1 Object-oriented data model

Il modello object-oriented rappresenta i dati in modo molto simile al sistema di tipi di un linguaggio ad oggetti. Esso è stato introdotto alla fine degli anni '90, e si è anche definito un linguaggio di interrogazione standard, OQL, che però non è mai stato implementato nella sua interezza.

In pratica, questo modello non viene più usato (nella sua forma originale).

1.2 Object-relational data model

Questo modello si basa su quello relazionale, che viene esteso con alcune caratteristiche del modello a oggetti: gerarchie di tipi, tipi definiti dall'utente, metodi associati alle tabelle, ecc.

Lo standard di riferimento è SQL:1999 (chiamato anche SQL3), che è ampiamente implementato (anche se con alcune variazioni): tutti i principali DBMS "relazionali" moderni implementano in realtà il modello object-relational.

1.3 Tree-based data model

Il principale esempio di tree-based data model è *XML*, che è una serializzazione di un modello basato, appunto, su una rappresentazione ad albero. Ai fini della gestione dei dati rappresentati in XML, il W3C (World Wide Web Consortium) ha standardizzato alcuni linguaggi:

- il linguaggio di interrogazione *XQuery*, e il suo sottoinsieme *XPath*;
- il linguaggio *XML Schema*, per la definizione dello schema dei dati.

Nell'ambito delle basi di dati, XML non ha soppiantato il modello relazionale, ma, invece, è stato integrato in SQL (SQL/XML), mediante l'introduzione della possibilità di inserire interi alberi XML nelle celle di una tabella, e di eseguire interrogazioni su di essi mediante XQuery, oppure attraverso delle estensioni del linguaggio di query SQL.

Invece, XML è stato ampiamente utilizzato come formato di interscambio dei dati. A causa della sua verbosità, però, anche l'uso in questo ambito è diminuito, a favore di linguaggi con una sintassi più "leggera". Tra questi, il principale è JSON, che, come XML, è la serializzazione di un modello semi-strutturato ad albero, e anch'esso sta venendo incluso in SQL.

1.4 Graph-based data model

Con questi modelli, i dati vengono rappresentati e interrogati come grafi (composti da nodi e archi). Essi sono efficaci soprattutto per la rappresentazione di grandi quantità di dati che si presentano naturalmente come reti (ad esempio, le reti sociali, le interazioni tra le proteine, i contenuti di Wikipedia e la loro evoluzione nel tempo, ecc.).

Esistono vari approcci alla realizzazione di questo tipo di modelli:

RDF (Resource Data Format): un insieme di concetti che permettono di rappresentare grafi. Esso è standardizzato dal W3C, che ha definito anche un apposito linguaggio di interrogazione, SPARQL.

Un'applicazione di RDF è l'idea del "web semantico", cioè di aumentare le capacità di interrogazione dei motori di ricerca "attaccando" a ogni pagina web delle informazioni relative al suo significato, ai suoi contenuti.

Property graph: un grafo in cui i nodi e gli archi possono essere etichettati con altri dati. Per questo modello esistono diversi linguaggi d'interrogazione, tra cui Cypher (specifico per il DBMS Neo4j) e le iniziative di standardizzazione G-Core e GQL. Questi linguaggi sono piuttosto complicati, dato che interrogare un grafo è relativamente complesso (soprattutto rispetto a una tabella).

Knowledge graph: un'evoluzione dei property graph. Vengono usati, in generale, per rappresentare nel modo più ricco possibile l'informazione (ad esempio, Google li impiega per generare i riassunti che vengono presentati quando si ricercano personaggi famosi).