

Laboratorio Basi di Dati

Anno accademico 2023/2024

Bortuzzo Francesco

francesco.bortuzzo@spes.uniud.it 157430

Indice

1. Introduzione e obbiettivo del progetto	1
2. Confronto tra modelli di basi di dati	2
2.1. Basi di dati relazionali	2
2.2. Basi di dati a grafo	2
3. Neo4j	2
3.1. Specifica di implementazione dei nodi	2
3.2. Specifica di implementazione delle relazioni	2
4. Legionella	2
4.1. Introduzione al batterio	2
4.2. Requisiti non strutturati	2
4.3. Requisiti strutturati	2
4.4. Glossario	2
5. implementazione su base di dati a grafo	2
5.1. descrizione della struttura dei nodi	3
5.2. descrizione delle relazioni tra nodi	3
5.3. implementazione	3
5.4. operazioni	3
5.5. popolamento	3
6. Soluzione tradizionale vs soluzione a grafo	3
7. grafici	3
8. bibliografia	3
9. appunti	3

1. Introduzione e obbiettivo del progetto

La legionella è un batterio che si sviluppa in ambienti acquatici e umidi. La sua diffusione è particolarmente pericolosa per la salute umana, in quanto può causare la legionellosi, una malattia polmonare che può portare alla morte. La legionellosi è una malattia infettiva che si manifesta con sintomi simili a quelli dell'influenza, come febbre, tosse, dolori muscolari e mal di testa. La malattia può essere contratta inalando aerosol contenenti il batterio, come ad esempio le goccioline d'acqua presenti negli impianti di condizionamento e nei sistemi di riscaldamento. È pertanto di fondamentale importanza monitorare la diffusione di tale batterio negli ambienti umidi e acquatici, al fine di prevenire la diffusione della malattia e proteggere la salute pubblica. Centri di particolare interesse sono le strutture ospedaliere e le strutture termali/alberghiere che, per la loro natura, sono ambienti a rischio di diffusione del batterio.

La raccolta di dati relativi alla presenza del batterio nel nostro Paese è attuata da diversi enti e istituzioni. In tal senso un contributo si deve ai diversi organismi che costituiscono

il SNPA¹, di cui fa parte l'ARPA FVG² e il Ministero della Salute. I dati raccolti sono utilizzati per monitorare la diffusione del batterio e adottare misure di prevenzione e controllo.

Al fine di analizzare i dati raccolti e studiare la diffusione del batterio, è opportuno utilizzare un sistema informativo che permetta di memorizzare, gestire e interrogare i dati in modo efficiente. Tuttavia, in regione, il vasto numero di dati già raccolti non è stato organizzato in modo efficace e pertanto non è possibile effettuare ricerche senza prima ristrutturare e connettere i vari dataset.

In questo contesto, i sistemi di basi di dati giocano un ruolo fondamentale, in quanto permettono di memorizzare grandi quantità di dati e di effettuare ricerche complesse in modo rapido ed efficiente. In particolare, i sistemi di basi di dati a grafo sembrano particolarmente adatti per la modellazione e l'analisi di dati complessi, come quelli relativi alla diffusione della legionella poichè permettono di rappresentare le relazioni tra i dati in modo naturale e di effettuare ricerche complesse in modo efficiente.

Questo documento si propone di condurre un'analisi critica di un database relazionale nel contesto precedentemente delineato e di proporre un'alternativa tramite l'utilizzo di un database a grafo. Nello specifico, verrà illustrato il procedimento di modellazione e di creazione/popoloamento di un database a grafo utilizzando Neo4j, per l'analisi dei dati riguardanti la diffusione della legionella nella nostra regione.

2. Confronto tra modelli di basi di dati

2.1. Basi di dati relazionali

2.2. Basi di dati a grafo

3. Neo4j

3.1. Specifica di implementazione dei nodi

3.2. Specifica di implementazione delle relazioni

(cypher)

4. Legionella

4.1. Introduzione al batterio

4.2. Requisiti non strutturati

4.3. Requisiti strutturati

4.4. Glossario

5. implementazione su base di dati a grafo

¹Sistema nazionale per la protezione dell'ambiente

²Agenzia Regionale per la Protezione dell'Ambiente Friuli Venezia Giulia

5.1. descrizione della struttura dei nodi

5.2. descrizione delle relazioni tra nodi

(introduzione di uno schema generale con nodi e relazioni tra essi)

5.3. implementazione

5.4. operazioni

5.5. popolamento

6. Soluzione tradizionale vs soluzione a grafo

confronto in termini di spazio occupato

confronto tra complessità delle query

confronto dei tempi

7. grafici

8. bibliografia

9. appunti

sono reperibili i file sorgente del db relazionale? Se no lo implemento io?

utilizzo la stessa notazione iden1x per la sezione dedicata al db relazionale

come popolare il db relazionale

genero qualche grafico per studiare i dati del db o faccio il confronto solamente tra gli aspetti principali dei due modelli