

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. Introduzione al batterio Legionella	2
3. Confronto tra modelli di basi di dati	2
3.1. Basi di dati relazionali	2
3.2. Basi di dati a grafo	2
4. Neo4j	3
4.1. Specifica di implementazione dei nodi	3
4.2. Specifica di implementazione delle relazioni	3
5. Legionella	3
5.1. Introduzione al batterio	3
5.2. Requisiti non strutturati	3
5.3. Requisiti strutturati	3
5.4. Glossario	3
6. implementazione su base di dati a grafo	3
6.1. descrizione della struttura dei nodi	3
6.2. descrizione delle relazioni tra nodi	3
6.3. implementazione	3
6.4. operazioni	3
6.5. popolamento	3
7. Soluzione tradizionale vs soluzione a grafo	3
8. grafici	3
9. bibliografia	3
10. appunti	3

1. Introduzione e obbiettivo del progetto

Il batterio Legionella è un bacillo gram-negativo aerobio, non mobile, che si sviluppa in ambienti acquatici e umidi sia naturali, come acque sorgive, termali, di fiumi o di laghi, che artificiali, come tubature, serbatoi, fontane e piscine.

Il genere comprende 62 specie diverse, suddivise in 71 sierotipi, di cui circa 20 sono patogeni per l'uomo. La specie più comune è la Legionella pneumophila, responsabile della maggior parte dei casi di legionellosi.

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. La malattia può manifestarsi in due forme: la forma più comune è la legionellosi non polmonare, che si manifesta con sintomi influenzali e può essere facilmente curata con antibiotici; la forma

più grave è la legionellosi polmonare, che si manifesta con sintomi simili a quelli della polmonite e può portare a complicazioni gravi, come la polmonite atipica o il decesso. Il primo episodio registrato, da cui deriva il nome del batterio, risale al 1976 quando un'epidemia colpì i partecipanti di un raduno della Legione Americana in un famoso hotel di Philadelphia. In quell'occasione, 224 persone contrassero una forma di polmonite al tempo sconosciuta, risultata fatale per 32 di esse. Le indagini effettuate in tale occasione identificarono nel sistema di aria condizionata dell'albergo il mezzo di propagazione del batterio.

È dunque di fondamentale importanza monitorare la diffusione di questo batterio negli ambienti umidi e acquatici, al fine di prevenire la malattia e proteggere la salute pubblica. Particolare attenzione deve essere rivolta alle strutture ospedaliere e alle strutture termali e alberghiere, che per loro natura rappresentano 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 significativo proviene dai vari organismi che costituiscono il SNPA¹, di cui fa parte l'ARPA FVG² e il Ministero della Salute. Le informazioni raccolte sono utilizzate per monitorare la diffusione del batterio e adottare misure di prevenzione e controllo.

Al fine di analizzare i dati acquisiti 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 mira a condurre un'analisi critica di un database relazionale nell'ambito delineato e a proporre un'alternativa attraverso l'impiego di un database a grafo. In particolare, si illustrerà il processo di modellazione, creazione e popolamento di un database a grafo utilizzando Neo4j per l'analisi dei dati sulla diffusione della legionella nella nostra regione.

2. Introduzione al batterio Legionella

3. Confronto tra modelli di basi di dati

3.1. Basi di dati relazionali

3.2. Basi di dati a grafo

¹Sistema Nazionale per la Protezione dell'Ambiente

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

4. Neo4j

4.1. Specifica di implementazione dei nodi

4.2. Specifica di implementazione delle relazioni (cypher)

5. Legionella

5.1. Introduzione al batterio

5.2. Requisiti non strutturati

5.3. Requisiti strutturati

5.4. Glossario

6. implementazione su base di dati a grafo

6.1. descrizione della struttura dei nodi

6.2. descrizione delle relazioni tra nodi (introduzione di uno schema generale con nodi e relazioni tra essi)

6.3. implementazione

6.4. operazioni

6.5. popolamento

7. Soluzione tradizionale vs soluzione a grafo

confronto in termini di spazio occupato

confronto tra complessità delle query

confronto dei tempi

8. grafici

9. bibliografia

10. 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