

Laboratorio Basi di Dati

Anno accademico 2023/2024

Bortuzzo Francesco

francesco.bortuzzo@spes.uniud.it 157430

Indice

| | |
|---|---|
| 1. Introduzione e obiettivo del progetto | 2 |
| 1.1. Introduzione al batterio Legionella | 2 |
| 1.2. Legionella in Friuli Venezia Giulia | 2 |
| 1.3. Obiettivo del progetto | 3 |
| 2. Analisi dei requisiti | 4 |
| 2.1. Requisiti non strutturati | 4 |
| 2.2. Requisiti strutturati | 4 |
| 2.3. Glossario | 4 |
| 3. Confronto tra modelli di basi di dati | 4 |
| 3.1. Basi di dati relazionali | 4 |
| 3.2. Basi di dati a grafo | 4 |
| 4. Neo4j | 4 |
| 4.1. Specifica di implementazione dei nodi | 4 |
| 4.2. Specifica di implementazione delle relazioni | 4 |
| 5. implementazione su base di dati a grafo | 4 |
| 5.1. descrizione della struttura dei nodi | 4 |
| 5.2. descrizione delle relazioni tra nodi | 5 |
| 5.3. implementazione | 5 |
| 5.4. operazioni | 5 |
| 5.5. popolamento | 5 |
| 6. grafici | 5 |
| Bibliografia | 5 |
| 7. appunti | 5 |

1. Introduzione e obbiettivo del progetto

1.1. Introduzione al batterio Legionella

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. La Legionella è in grado di sopravvivere in una vasta gamma di condizioni ambientali, tra cui temperature comprese tra 20 e 45 gradi Celsius, pH neutro o leggermente alcalino e presenza di nutrienti organici. 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. 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 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.

1.2. Legionella in Friuli Venezia Giulia

Nell'Unione Europea, la raccolta di dati relativi alla presenza del batterio è effettuata dall'ECDC¹, un ente istituito nel 2005. Nel nostro Paese, invece, questa attività è svolta da diversi enti e istituzioni. Un contributo significativo proviene dal Ministero della Salute e dai vari organismi che costituiscono il SNPA², di cui fa parte l'ARPA FVG³. I dati raccolti sono utilizzati per valutare il rischio di diffusione del batterio e adottare misure di prevenzione e controllo.

Proprio l'ARPA FVG ha condotto diverse indagini sulla presenza di Legionella nei sistemi di aduzione e conservazione dell'acqua della regione e ha pubblicato i risultati in vari report. Ad esempio nel 2019 ARPA FVG e Università degli studi di Udine hanno collaborato per la pubblicazione di un articolo (Felice et al., 2019) riguardante la presenza di legionella nei sistemi di raccolta e distribuzione dell'acqua nella regione, in un periodo

¹Centro europeo per la prevenzione e il controllo delle malattie

²Sistema Nazionale per la Protezione dell'Ambiente

³Agenzia Regionale per la Protezione dell'Ambiente Friuli Venezia Giulia

di sedici anni, dal 2002 al 2017, durante il quale sono stati raccolti e analizzati 20.319 campioni in 3.983 indagini ambientali.

I risultati derivati dallo studio, riferiti alle indagini non cliniche⁴, hanno evidenziato che la presenza di Legionella è diffusa soprattutto nei cluster di impianti termali (il batterio è stato individuato nel 57,8% dei siti indagati) e ospedalieri (nel 50,8% delle strutture è stata riscontrata la legionella almeno una volta) con picchi dei campioni positivi soprattutto nei mesi che segnano l'inizio del periodo autunnale. Sebbene la presenza del batterio abbia mostrato un trend crescente nel periodo considerato, si è osservato un forte incremento tra la seconda metà del 2006 e l'inizio del 2009, seguito da un calo fino al 2013 e un nuovo aumento negli anni successivi. Questo andamento evidenzia la necessità di monitorare costantemente la presenza di Legionella e di adottare misure di prevenzione e controllo per evitare la diffusione del batterio e, conseguentemente, ridurre il rischio di nuove epidemie.

1.3. Obiettivo del progetto

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.

⁴eseguite routinariamente nell'ambito del piano regionale di sorveglianza ambientale

2. Analisi dei requisiti

2.1. Requisiti non strutturati

Prima di poter procedere con la modellazione e l'implementazione del database, è necessario definire i requisiti non strutturati del sistema informativo. Questi requisiti sono di natura qualitativa e descrivono le caratteristiche che il sistema deve possedere per soddisfare le esigenze degli utenti e degli stakeholder. Nel nostro ambito i requisiti non strutturati riguardano l'intera fase di acquisizione dei dati relativi alla legionella. In particolare, i requisiti non strutturati del sistema informativo sono i seguenti:

Il sistema deve permettere la registrazione di indagini ambientali relative alla presenza di Legionella nei sistemi di aduzione e conservazione dell'acqua.

Un'indagine ambientale consiste nel prelievo di campioni in un sito specifico, in una data stabilita, per analizzarli alla ricerca della presenza di Legionella.

Ognuno dei campioni prelevati deve essere associato a un sito specifico, a una data di prelievo e ad un codice identificativo univoco. Tutti i campioni prelevati devono essere sottoposti a diverse analisi per determinare la presenza o l'assenza di Legionella:

1. PCR⁵ qualitativa: permette di identificare la presenza del DNA di Legionella nei campioni prelevati.
2. PCR quantitativa: permette di quantificare il numero di copie di DNA di Legionella presenti nei campioni prelevati.
3. Coltura: permette di isolare e identificare le colonie di Legionella presenti nei campioni prelevati.

I risultati delle analisi devono essere registrati nel sistema informativo e associati ai campioni prelevati.

2.2. Requisiti strutturati

2.3. Glossario

3. Confronto tra modelli di basi di dati

3.1. Basi di dati relazionali

3.2. Basi di dati a grafo

4. Neo4j

4.1. Specifica di implementazione dei nodi

4.2. Specifica di implementazione delle relazioni (cypher)

5. implementazione su base di dati a grafo

5.1. descrizione della struttura dei nodi

⁵Polymerase Chain Reaction, tecnica che consiste nell'amplificazione dei frammenti di acidi nucleici

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. grafici

Bibliografia

Felice, A., Franchi, M., De Martin, S., Vitacolonna, N., Iacumin, L., & Civilini, M. (2019). Environmental surveillance and spatio-temporal analysis of *Legionella* spp. in a region of northeastern Italy (2002–2017). *PLOS ONE*, 14(7), e218687. <https://doi.org/10.1371/journal.pone.0218687>

7. appunti

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