

Data Science e Tecnologie per le Basi di Dati

Quaderno 4 – MongoDB

Introduzione

Viene fornito il dataset “bike stations”, contenente informazioni circa 65 stazioni di un servizio di bike sharing. Per lo svolgimento del quaderno, è consigliato importare il dataset come collection su un database su server MongoDB locale, utilizzando il tool `mongoimport`.

Nota: Per l’installazione di MongoDB in locale, è necessario fare riferimento a “MongoDB Community Edition”, installabile seguendo la [documentazione ufficiale](#).

Nota: Per utilizzare il tool `mongoimport`, è necessario installare “MongoDB Database Tools”, [MongoDB Database Tools](#)).

È possibile fare riferimento alle indicazioni fornite nel laboratorio su MongoDB per i comandi di importazione di una collection, di esecuzione di un server MongoDB in locale e di accesso alla command-line interface (shell) MongoDB.

Di seguito viene riportato un esempio di stazione estratta dalla collection.

```
{
  "_id" : ObjectId("61b75b13fd4d2d1ea82e75f4"),
  "empty_slots" : 10,
  "extra" : {
    "number" : 57,
    "reviews" : 222,
    "score" : 4,
    "status" : "online",
    "uid" : "307"
  },
  "free_bikes" : 4,
  "id" : "bfa12cb895ac0d7392dde60b6b433cdf",
  "name" : "San Francesco da Paola",
  "timestamp" : "2021-12-10T14:54:39.185000Z",
  "location" : {
    "type" : "Point",
    "coordinates" : [
      45.068617,
      7.689097
    ]
  }
}
```

Per rispondere alle domande, è necessario riportare:

- L’interrogazione utilizzata per ottenere la risposta alla domanda (la query deve estrarre solo i campi necessari a rispondere alla domanda)
- La risposta alla domanda

Domanda 1

Quante stazioni hanno rispettivamente status (extra.status) "online" e "offline"?

Domanda 2

Quante stazioni hanno uno status diverso da "online" e "offline"?

Domanda 3

Per le stazioni che hanno uno status diverso da "offline" e "online", visualizzare solo il valore del campo status.

Domanda 4

Quali sono le stazioni attive (status = online) con una valutazione media (extra.score) maggiore o uguale a 4?

Estrarre l'elenco dei nomi di queste stazioni, ordinato in ordine alfabetico.

Domanda 5

Qual è il nome delle stazioni non attive (status = offline) che hanno almeno una postazione libera (empty_slots > 0) oppure hanno almeno una bici a disposizione (free_bikes > 0)? Quante postazioni libere e quante bici sono a disposizione?

Domanda 6

Qual è il numero totale di recensioni (extra.reviews) per tutte le stazioni?

Domanda 7

Per ciascun valore di valutazioni medie (score), quante sono le stazioni a cui è stata assegnata quella valutazione? Ordinare il risultato per valutazione decrescente.

Domanda 8

Qual è la valutazione media per le stazioni attive (status = online) e non attive (status = offline)?

Nota: le stazioni che non rientrano in nessuna delle due categorie (vedi domanda 3) non andranno considerate nella query di conteggio.

Domanda 9

Quali sono le valutazioni medie per le stazioni senza bici (free_bikes = 0) e per quelle con almeno una bici a disposizione (free_bikes > 0)?

Suggerimento: È possibile utilizzare il paradigma [map-reduce](#) per rispondere a questa domanda. La funzione mapReduce() è stata deprecata in MongoDB 5.0. Tuttavia, il paradigma resta una valida alternativa, utilizzata, per esempio, in Hadoop. Per questo motivo, ne viene consigliato l'utilizzo per la risoluzione di questo esercizio.

Domanda 10

Rispondere alla domanda 9, facendo riferimento solamente alle stazioni attive (status = online).

Suggerimento: Anche per questo esercizio, è consigliato l'utilizzo del paradigma map-reduce.

Domanda 11

Quali sono i nomi delle 3 stazioni con bici disponibili (`free_bikes > 0`) più vicine al punto `[45.07456, 7.69463]`?
Quante bici sono disponibili?

Nota: È necessario creare un indice `2dsphere` su `"location"` per utilizzare l'operatore `$near`.

Nota: È possibile usare il metodo `limit(n)` per limitare il numero di risultati estratti.

Domanda 12

Quali sono i nomi delle 3 stazioni con bici disponibili (`free_bikes > 0`) più vicine alla stazione `"Politecnico 4"`?
Quante bici sono disponibili?

Nota: È necessario creare un indice `2dsphere` su `"location"` per utilizzare l'operatore `$near`.

Requisito: Risolvere l'esercizio utilizzando una query annidata per l'estrazione della posizione della stazione `"Politecnico 4"`.