

Business Intelligence per i Big Data

Esercitazione n. 10

Parte 1 – Compass

L'obiettivo di questa prima parte del laboratorio è quello di utilizzare MongoDB Compass come tool visuale per lavorare con un db non relazionale. In questo esercizio vi verrà richiesto di analizzare i dati e scrivere alcune query per prelevare informazioni da un database NoSQL basato su MongoDB.

1) Setup e connessione al database remoto

Installazione di MongoDB Compass

1. Download MongoDB Compass a <https://www.mongodb.com/try/download/compass>
2. Installa e apri l'applicazione.
3. Crea un cluster gratuito (se non ne hai già uno)
 - a. Crea un account MongoDB (<https://www.mongodb.com/cloud/atlas/register>)
 - b. Seleziona "Shared Cluster" (free cluster). Le impostazioni di default sono quelle per ottenere un account completamente gratuito.
 - c. Dai un nome al tuo cluster.
 - d. Vai su Accesso al database
 - i. Fare clic su "Aggiungi nuovo utente del database"
 - ii. Selezionare il metodo di autenticazione con password
 - iii. Compila il modulo
 - iv. Lasciare tutte le opzioni predefinite
 - e. Configurare l'accesso remoto
 - i. Clicca su "Accesso alla rete"
 - ii. Clicca su "Aggiungi indirizzo IP"
 - iii. Inserire nel campo di immissione dell'elenco di accesso 0.0.0.0/0
4. Ottieni la stringa di connessione dal server MongoDB Atlas
 - a. Vai alla pagina "Distribuzioni database" dal menu laterale
 - b. Fare clic su "Connetti" vicino al cluster appena creato
 - c. Seleziona "Compass"
 - d. Copiare la stringa di connessione
5. Incolla la stringa di connessione in MongoDB Compass
6. Fare clic su Connetti

Creazione di un database

7. Fare clic su Crea database
 - a. Assegnare un nome al database
 - b. Inserire il nome "Prenotazioni" nel campo "Nome Collezione"
 - c. Lasciare tutti i valori predefiniti per le altre opzioni
8. Aggiungi una nuova collezione
 - a. Fare clic sul pulsante "Crea collezione"
 - b. Inserisci il nome "Parcheggi"
 - c. Lasciare tutti i valori predefiniti per le altre opzioni
9. Scarica i dati della collezione dal sito web

10. Aggiungi dati a ciascuna collezione
 - a. Seleziona una collezione
 - b. Fare clic sul pulsante "Aggiungi dati", quindi Importa file
 - c. Selezionare il file della collezione corrispondente
 - d. Seleziona il tipo Json
 - e. Fare clic su Importa
 11. Seleziona la collezione Parcheggi
 12. Vai alla scheda "Indici"
- Creare un indice 2dsphere sul campo loc

Descrizione del Database Parkings

Il database Parkings contiene informazioni relative ai parcheggi di auto associate ad un servizio di Car Sharing. Gli attributi più importanti per ogni documento sono elencati nella tabella 1.

Tabella 1: descrizione del database Parkings.

Nome	Tipo	Descrizione
_id	objectid	Identificativo del documento
address	string	Indirizzo di parcheggio del veicolo
city	string	Città nella quale è avvenuto il parcheggio
engineType	string	Identificativo del tipo di motore del veicolo
exterior	string	Stringa identificativa delle condizioni esterne del veicolo
final_date	date	Data e ora di fine stazionamento del veicolo
fuel	int32	Livello di carburante del veicolo durante il periodo di stazionamento
init_date	date	Data e ora di inizio stazionamento del veicolo
interior	string	Stringa identificativa delle condizioni interne del veicolo
loc	coordinates	Coordinate della posizione di parcheggio del veicolo
plate	int32	Identificativo della targa del veicolo
smartphoneRequired	boolean	Valore binario (Vero/Falso) che indica qualora sia necessario lo smartphone per il noleggio del veicolo
vendor	string	Azienda proprietaria del veicolo da noleggiare
vin	string	Identificativo del telaio del veicolo

Descrizione del Database Bookings

Il database Bookings contiene informazioni relative al noleggio di veicoli per un servizio di Car Sharing. Gli attributi più importanti per ogni documento sono elencati nella tabella 2.

Tabella 2: descrizione del database Bookings.

Nome	Tipo	Descrizione
_id	objectid	Identificativo del documento
car_name	string	Modello del veicolo

Nome	Tipo	Descrizione
city	string	Città nella quale è avvenuto il noleggio
distance	int32	Distanza percorsa durante il noleggio (in metri)
driving	object (int32, int32)	Distanza percorsa durante il noleggio (in metri) e Durata del noleggio (in secondi)
engineType	string	Identificativo del tipo di motore del veicolo
exterior	string	Stringa identificativa delle condizioni esterne del veicolo
final_address	string	Indirizzo della posizione di fine noleggio
final_date	date	Data e ora di fine noleggio
final_fuel	int32	Livello del carburante a fine noleggio
init_address	int32	Indirizzo della posizione di inizio noleggio
init_date	date	Data e ora di inizio noleggio
init_fuel	int32	Livello del carburante a inizio noleggio
interior	string	Stringa identificativa delle condizioni interne del veicolo
plate	int32	Identificativo della targa del veicolo
smartphoneRequired	boolean	Valore binario (Vero/Falso) che indica qualora sia necessario lo smartphone per il noleggio del veicolo
vendor	string	Azienda proprietaria del veicolo da noleggiare
walking	object (int32, int32)	Distanza percorsa a piedi per raggiungere il veicolo e Durata del percorso a piedi per raggiungere il veicolo

1. Analizzare la base dati con lo schema analyzer

Parkings database

1. Identificare l'intervallo/gli intervalli orari con maggiore richiesta di parcheggio (init_date) di veicoli.
2. Identificare l'intervallo/gli intervalli orari nei quali i veicoli vengono noleggiati (final_date) più di frequente.
3. Filtrare sulla mappa una zona di interesse (loc) e analizzare l'intervallo/gli intervalli orari di inizio noleggio (fine stazionamento) più frequenti.
4. Per i veicoli filtrati al passo precedente, visualizzare solo quelli che hanno un livello di carburante residuo maggiore del 90%.
5. Visualizzare su mappa i veicoli che hanno un livello di carburante residuo inferiore al 5%.

Bookings database

6. Identificare la/le percentuali più frequenti di livello di carburante a inizio noleggio.
7. Identificare la/le percentuali più frequenti di livello di carburante a fine noleggio.
8. Identificare il range di distanza più frequente se il percorso fosse stato fatto a piedi.

2. Interrogare la base dati (Parkings)

Scrivere le interrogazioni in Documents, aprire la finestra Options. Il primo campo è il filter. Per runnare le interrogazioni, utilizzare il bottone Find.

9. (Parkings) Trovare le targhe e gli indirizzi di parcheggio dei veicoli che hanno iniziato il noleggio (finito stazionamento) dopo le 6.00 del 30-09-2017.

Hint: usare la funzione:

```
Date("<YYYY-mm-ddTHH:MM:ss>")
```

10. (Parkings) Trovare gli indirizzi e il livello di carburante residuo per le auto che hanno avuto durante lo stazionamento almeno il 70% di carburante residuo e ordinare i risultati in base al loro livello di carburante decrescente.
11. (Parkings) Trovare la targa, tipo di motore, e livello di carburante dei veicoli di 'car2go' che hanno buone condizioni interne ed esterne.

3. Interrogare la base dati (Bookings)

12. (Bookings) Per i noleggi che hanno un percorso a piedi maggiore di 15 km, visualizzare la data e l'orario di inizio noleggio e il livello di carburante a inizio noleggio. Visualizzare i risultati ordinati in base al livello di carburante iniziale decrescente.
13. (Bookings) Raggruppare i documenti in base al loro livello di carburante a fine noleggio. Per ogni gruppo visualizzare il livello di carburante medio a inizio noleggio.
14. (Bookings) Visualizzare la distanza media percorsa nei noleggi per ciascun fornitore del servizio. In media con quale fornitore del servizio gli utenti percorrono una distanza maggiore?
15. (Bookings) Quali sono i primi tre veicoli che hanno percorso una maggiore distanza considerando tutti i noleggi? Visualizzarne la targa e il conteggio in ordine decrescente per totale distanza percorsa in tutti i noleggi.
16. (Bookings) Quanti sono i veicoli che, considerando tutti i noleggi, hanno percorso almeno 350000m?
17. (Bookings) Ripetere la query precedente considerando solamente i veicoli di proprietà 'car2go'.

Interrogazioni Bonus

18. (Parkings) Quanti sono i parcheggi effettuati a meno di 1km da Piazza San Carlo (coordinate 7.683016, 45.067764)?

(Il raggio della Terra espresso in km è pari a 6378.1 km.)

Hint: usare l'operatore \$geoWithin insieme a \$centerSphere:

```
$geoWithin: {$centerSphere: [[7.683016, 45.067764], 1/6378.1]}
```

19. Ripetere l'interrogazione al passo precedente con un punto di interesse personale nell'area metropolitana di Torino (e.g. indirizzo di casa) usando Open Street Maps per trovare le coordinate esatte (www.openstreetmap.org, invertire l'ordine delle coordinate).