

Business Intelligence per i Big Data

Esercitazione n. 10

Passi preliminari per lo svolgimento dell'esercitazione

L'obiettivo di questa esercitazione è quello di prendere confidenza con la struttura dei database NoSQL, esplorando e interrogando una base dati grazie al tool MongoDB Compass.

Connessione alla base dati remota

1. Scaricare MongoDB Compass dal seguente link:
<https://www.mongodb.com/try/download/compass>
Scegliere la versione corrispondente al sistema operativo utilizzato e, al termine del download, lanciare l'eseguibile.
2. Connettersi alla base dati remota tramite il seguente URI:
mongodb://bibd_student:bibd_student@mp1.polito.it:8005/test
3. Click su **Connect**.
4. Accedere su **bibd**.
5. Accedere alla base **Parkings o Bookings**.

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.

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

Tabella 1: descrizione del database Parkings.

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.

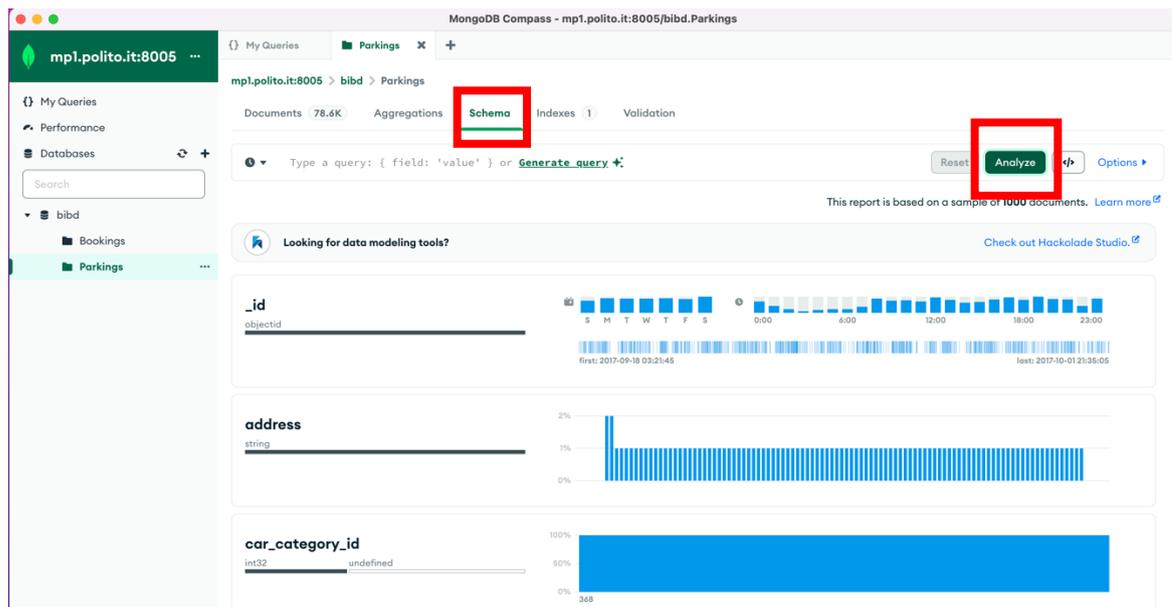
Nome	Tipo	Descrizione
_id	objectid	identificativo del documento
car_name	string	Modello del veicolo
city	string	Città nella quale è avvenuto il noleggio
distance	int32	Distanza percorsa durante il noleggio (in metri)
driving (int32, int32)	object	Distanza percorsa durante il noleggio (in metri) e Durata 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 (int32, int32)	object	Distanza percorsa a piedi per raggiungere il veicolo e Durata del percorso a piedi per raggiungere il veicolo

Tabella 2: descrizione del database Bookings.

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%.

2. Concentrarsi ora sul database Bookings.

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

3. Interrogare la base dati (Parkings)

The screenshot shows the MongoDB Compass interface for the 'Parkings' collection. The 'Documents' tab is active, displaying a list of documents. The first document is highlighted, showing fields such as '_id', 'city', 'vendor', 'final_time', 'plate', 'car_category_id', 'init_time', 'virtual_rental_type_id', 'car_model_data', 'car_name', 'init_date', 'on_click_disabled', 'virtual_rental_id', 'fuel', 'final_date', 'loc', and 'address'. The 'Options' button is highlighted in red. Below the main view, a query editor is shown with fields for Project, Sort, Collation, and Index Hint, along with Max Time MS, Skip, and Limit options.

Scrivere le interrogazioni in *Documents*, aprire la finestra *Options*. Il primo campo è il filter.

Per runnare le interrogazioni, utilizzare il bottone *Find*.

1. (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>")`)
2. (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.
3. (Parkings) Trovare la targa, tipo di motore, e livello di carburante dei veicoli di 'car2go' che hanno buone condizioni interne ed esterne.

4. Interrogare la base dati (Bookings)

1. (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.
2. (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.
3. (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?
4. (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.
5. (Bookings) Quanti sono i veicoli che, considerando tutti i noleggi, hanno percorso almeno 350000m?
6. (Bookings) Ripetere la query precedente considerando solamente i veicoli di proprietà 'car2go'.

Interrogazioni Bonus

1. (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](#) in questo modo:
\$geoWithin: {\$centerSphere: [[7.683016, 45.067764], 1/6378.1]}
2. 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).