

Business Intelligence per i Big Data

Esercitazione di laboratorio n. 7

Passi preliminari per lo svolgimento delle esercitazioni

L'obiettivo di questa esercitazione è quello di prendere confidenza con la struttura dei database NoSQL. Questa prima esercitazione tratterà l'esplorazione e l'interrogazione della base dati grazie al tool MongoDB Compass.

Connessione alla base dati remota

1. Posizionarsi su c:\mongodb-compass e lanciare MongoDB Compass tramite l'apposita icona.
Qualora il file .exe non fosse disponibile scaricarlo all'indirizzo: <https://downloads.mongodb.com/compass/mongodb-compass-1.18.0-win32-x64.exe> (versione Windows)
2. Connettersi alla base dati remota tramite le seguenti credenziali:
 - a. **Hostname:** bigdatadb.polito.it
 - b. **Port:** 27017
 - c. **Authentication:** Username/Password
 - d. **Username:** Compass
 - e. **Password:** Compass19!
 - f. **Authentication database:** dbdmg
 - g. **SSL:** Unvalidated (insecure)
3. (Opzionale) Specificare un **Favorite Name** per connettersi rapidamente alla base dati in futuro. Click su **Save Favorite**.
4. Click su **Connect**.
5. Accedere su **dbdmg**.
6. Accedere alla base **Parkings**.

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.

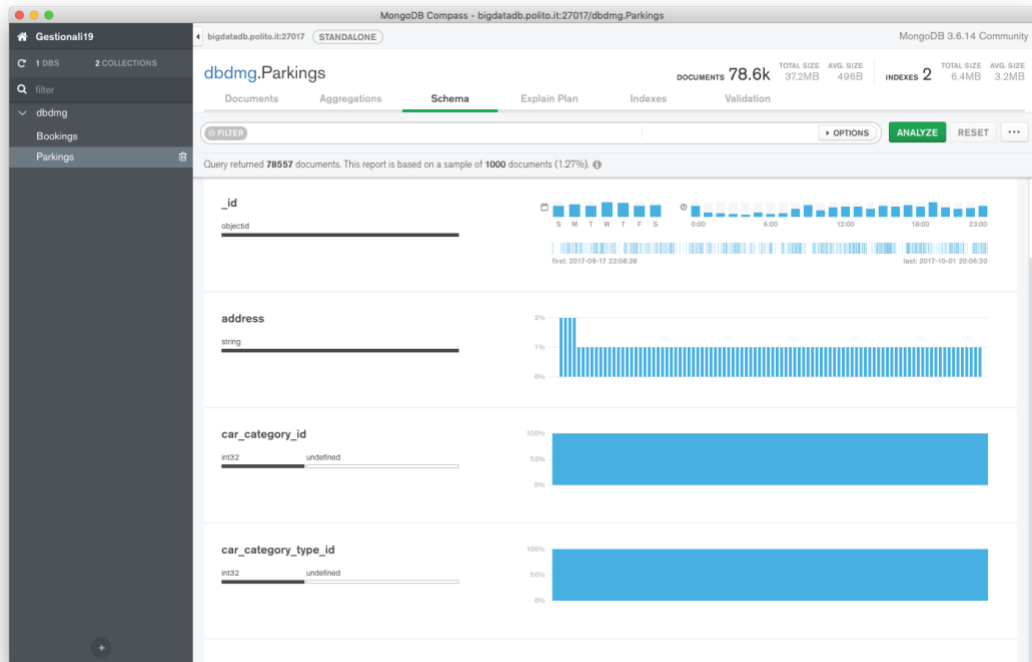
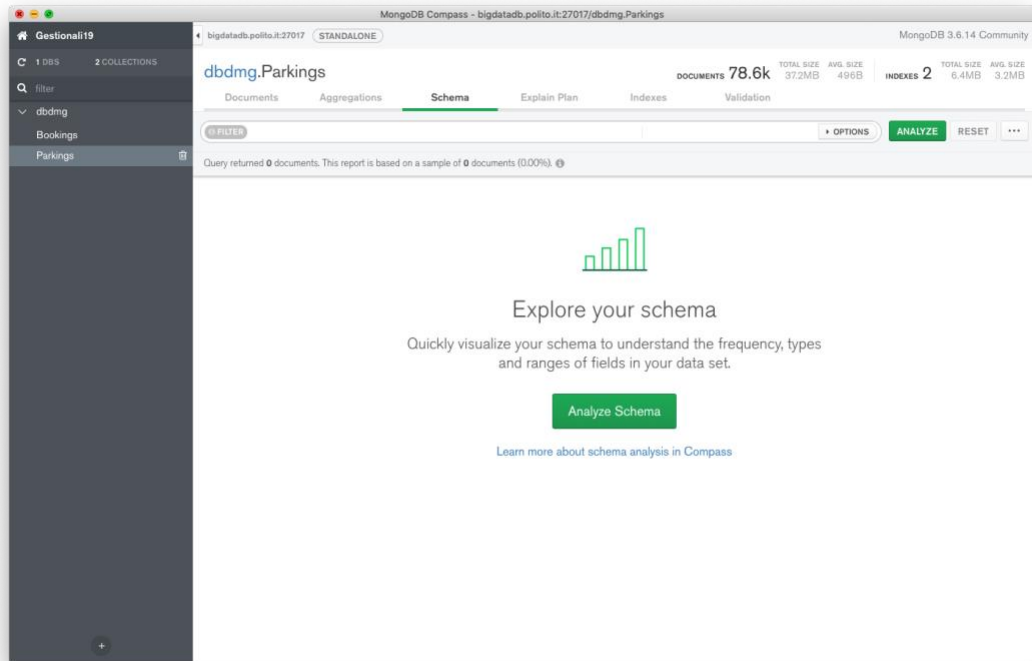
Descrizione del Database *Bookings*

Il database Parkings 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	object	distance	int32	Distanza percorsa durante il noleggio (in metri)
		duration	int32	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	distance	int32	Distanza percorsa a piedi per raggiungere il veicolo.
		duration	int32	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)

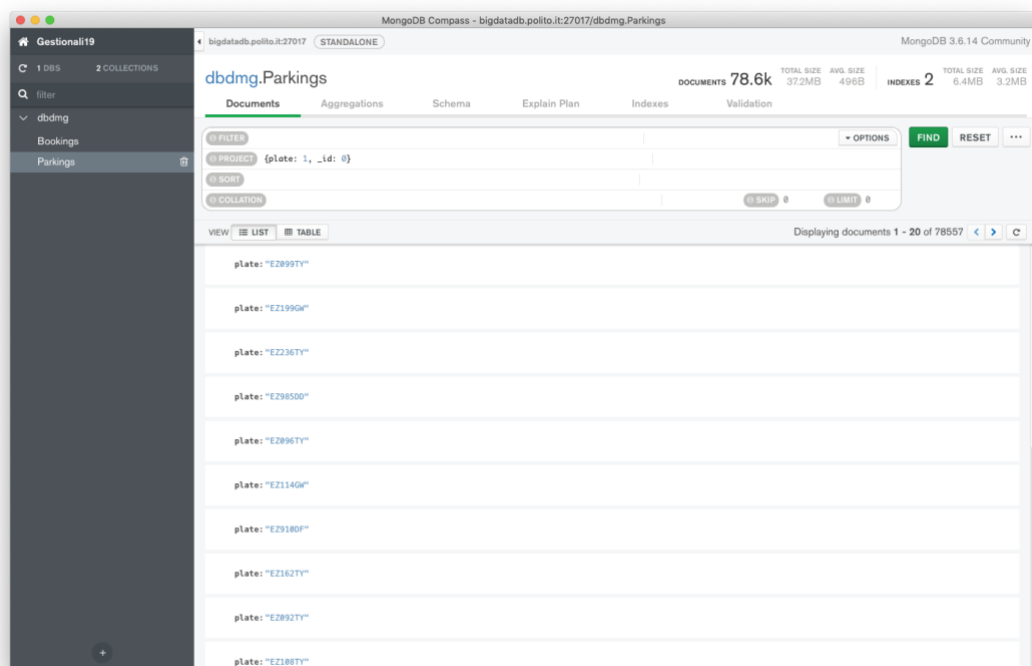


1. Identificare l'intervallo/gli intervalli orari con maggiore **richiesta di parcheggio (inizio stazionamento)** di veicoli.
2. Identificare l'intervallo/gli intervalli orari nei quali i veicoli **vengono noleggiati (fine stazionamento)** più di frequente.
3. Filtrare sulla mappa una zona di interesse 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%.

Concentrarsi ora sul database *Bookings*.

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)



1. 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. 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. Trovare la targa, tipo di motore, e livello di carburante dei veicoli di 'car2go' che hanno buone condizioni interne ed esterne.

4. (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.
5. (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.
6. (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?

Interrogazioni Bonus

7. (Parkings) Trovare le macchine parcheggiate a meno di 1 km da Piazza San Carlo (coordinate 7.683016, 45.067764).
Hint: usare l'operatore [\\$geoWithin](#) insieme a [\\$centerSphere](#).
8. (Parkings) 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).