

# Business Intelligence per i Big Data

---

## Quaderno #2 – MongoDB

### Introduzione

L'obiettivo del quaderno è quello di analizzare e sintetizzare informazioni utili da un database Mongo. Lo studio è condotto su un dataset reale che raccoglie le informazioni sui parcheggi e prenotazioni di veicoli per il *Car Sharing*.

### Dataset

L'analisi verte su dei dati reali relativi al servizio di Car Sharing nell'area metropolitana di Torino. Il dataset originale contiene ~160.000 documenti, divisi in due collezioni. Lo schema delle due collezioni è descritto di seguito.

Per accedere al database utilizzare le seguenti credenziali su MongoDB Compass:

- Hostname: bigdatadb.polito.it
- Port: 27017
- Authentication: Username/Password
- Username: Compass
- Password: Compass19!
- Authentication database: dbdmg
- SSL: Unvalidated (insecure)

### 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
<b>_id</b>	objectid	identificativo del documento
<b>address</b>	string	Indirizzo di parcheggio del veicolo
<b>city</b>	string	Città nella quale è avvenuto il parcheggio
<b>engineType</b>	string	Identificativo del tipo di motore del veicolo
<b>exterior</b>	string	Stringa identificativa delle condizioni esterne del veicolo
<b>final_date</b>	date	Data e ora di fine stazionamento del veicolo
<b>fuel</b>	int32	Livello di carburante del veicolo durante il periodo di stazionamento

<b>init_date</b>	date	Data e ora di inizio stazionamento del veicolo
<b>interior</b>	string	Stringa identificativa delle condizioni interne del veicolo
<b>loc</b>	coordinates	Coordinate della posizione di parcheggio del veicolo
<b>plate</b>	int32	Identificativo della targa del veicolo
<b>smartphoneRequired</b>	boolean	Valore binario (Vero/Falso) che indica qualora sia necessario lo smartphone per il noleggio del veicolo
<b>vendor</b>	string	Azienda proprietaria del veicolo da noleggiare
<b>vin</b>	string	Identificativo del telaio del veicolo

Tabella 1: descrizione del database Parkings.

## 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
<b>_id</b>	objectid	identificativo del documento
<b>car_name</b>	string	Modello del veicolo
<b>city</b>	string	Città nella quale è avvenuto il noleggio
<b>distance</b>	int32	Distanza percorsa durante il noleggio (in metri)
<b>driving</b>	object	<b>distance</b> int32    Distanza percorsa durante il noleggio (in metri)
		<b>duration</b> int32    Durata del noleggio (in secondi)
<b>engineType</b>	string	Identificativo del tipo di motore del veicolo
<b>exterior</b>	string	Stringa identificativa delle condizioni esterne del veicolo
<b>final_address</b>	string	Indirizzo della posizione di fine noleggio
<b>final_date</b>	date	Data e ora di fine noleggio
<b>final_fuel</b>	int32	Livello del carburante a fine noleggio
<b>init_address</b>	int32	Indirizzo della posizione di inizio noleggio
<b>init_date</b>	date	Data e ora di inizio noleggio
<b>init_fuel</b>	int32	Livello del carburante a inizio noleggio
<b>interior</b>	string	Stringa identificativa delle condizioni interne del veicolo
<b>plate</b>	int32	Identificativo della targa del veicolo

<b>smartphoneRequired</b>	boolean	Valore binario (Vero/Falso) che indica qualora sia necessario lo smartphone per il noleggio del veicolo		
<b>vendor</b>	string	Azienda proprietaria del veicolo da noleggiare		
<b>walking</b>	object	<b>distance</b>	int32	Distanza percorsa a piedi per raggiungere il veicolo.
		<b>duration</b>	int32	Durata del percorso a piedi per raggiungere il veicolo.

Tabella 2: descrizione del database Bookings.

## Specifiche per la generazione del report

Preparare una breve relazione (max 2 pagine) per descrivere i risultati dell'analisi includendo:

- la risposta a ciascuna domanda e, ove necessario, la query - o le query - usate per rispondere alla domanda, in formato testuale, secondo la sintassi di MongoDB;
- gli screenshot dei risultati ottenuti quando (gli screenshot non rientrano nel conteggio del numero massimo di pagine).

## Domande

Rispondere alle seguenti domande:

1. Quale periodo di tempo è coperto dalla base dati? Riportare uno screenshot che supporti la risposta.
2. Qual è il giorno della settimana in cui sono stati effettuati più noleggi? Riportare uno screenshot che supporti la risposta.
3. Qual è il l'intervallo orario della settimana in cui sono stati effettuati più noleggi? Riportare uno screenshot che supporti la risposta.
4. Quanti sono i parcheggi effettuati nel cerchio con centro nel punto A = [7.694652260765925, 45.07449636396572] e raggio R = 0.00010117446809534804?
5. Rispondere alle domande 2 e 3 considerando solo i parcheggi selezionati al punto 4. E' un risultato atteso?
6. Quali sono i veicoli che hanno percorso una maggiore distanza considerando tutti i noleggi? Visualizzarne la targa e il conteggio in ordine decrescente per conteggio.
7. Quali sono i fornitori distinti presenti nella base dati? Quanti noleggi sono stati registrati per ciascuno di essi?
8. Al punto 4 abbiamo analizzato l'attività nell'area in termini di parcheggi. Ora guardiamo l'affluenza in termine di noleggio che termina in quell'area. Scegliere un secondo punto nella mappa a piacimento (ad es. B = [7.695856, 45.064539]) lasciando R invariato. Quanti sono i viaggi effettuati che iniziano in B e terminano in A? (*Hint: le coordinate del luogo di partenza/arrivo del record si trovano in "origin\_destination.coordinates.0/1"*)
9. Rispondere alle domande 2 e 3 considerando solo i noleggi ottenuti al punto 8.
10. Se la base dati è consistente, il numero di noleggi che termina in A è uguale al numero di parcheggi rilevati in A. La base dati è consistente? Se no, provare a dare un'interpretazione del perché.