

Lista Attività A.A. 2022/2023 – MongoDB

Effettuare a scelta una delle seguenti attività e inviare via mail la consegna al docente (riccardo.martoglia@unimore.it)

1. Modellazione e interrogazione su dati reali

Partendo da un dataset (pubblico o creato da zero) non-JSON, convertirne in JSON la struttura modellandola in modo opportuno (attraverso uno script python, è possibile considerare anche solo una porzione dei campi originali). Quindi, inserire i dati in una collezione e interrogarli creando almeno 5 interrogazioni di crescente difficoltà. Tra queste dovranno essere incluse interrogazioni di aggregazione e che utilizzano alcuni comandi presenti nella documentazione non visti a lezione (il cui uso andrà commentato adeguatamente). Nota: l'aspetto di modellazione è fondamentale in questa attività, quindi sarà necessario descrivere con cura il dataset di partenza e dare una "forma" opportuna a quello creato, utilizzando e motivando tra l'altro gli strumenti / pattern di modellazione visti a lezione (es. liste, annidamenti, embedding, linking, ecc.).

Consegna: PDF commentato con dati di esempio, codice interrogazioni ed eventualmente codice Python per conversione (includere dati e codici anche in un file .txt per facilitarne il testing)

2. Indexing e query optimization o Text search

Approfondire un argomento tra questi studiando il capitolo 8 (Indexing e query optimization) o 9 (Text search) del libro "MongoDB in action" disponibile sul sito dell'insegnamento. Riassumere quindi alcune delle tecniche imparate e mostrarne un'applicazione pratica su alcune nuove interrogazioni sul dataset Amazon¹ utilizzato nelle esercitazioni (es. decidere gli indici adeguati ottimizzandone l'esecuzione o mostrare l'utilizzo di varie tecniche di text search, commentando le scelte effettuate).

Consegna: PDF commentato con breve relazione e codice interrogazioni / creazione degli indici (includere dati e codice anche in un file .txt per facilitarne il testing)

3. MongoDB Usecase e applicazione Python

Individuato un dataset JSON pubblico, possibilmente modificato e/o arricchito nei dati e nella loro struttura (giustificare le modifiche effettuate), scrivere il codice Python / PyMongo relativo ad un esempio di applicazione pratica nell'ambito dei dati scelti. In particolare, prevedere una funzione per caricare i dati su MongoDB, oltre che un insieme esemplificativo di funzioni che dimostrino alcune delle principali operazioni di gestione previste (particolare enfasi su interrogazioni anche complesse, ma anche con operazioni di scrittura/modifica). Per quanto riguarda le interrogazioni, dovranno essere utilizzati anche alcuni comandi presenti nella documentazione non visti a lezione (il cui uso andrà commentato adeguatamente). È anche possibile partire da un dataset creato da zero, che abbia una sufficiente complessità.

Consegna: PDF commentato con descrizione di dati e funzioni e relativo codice python (includere dati e codice anche in un file .txt per facilitarne il testing)

NOTE

¹ È possibile lavorare su una porzione a piacere del dataset, facendo riferimento al link pubblicato nelle slide di laboratorio o sfruttando direttamente il dataset presente nelle macchine virtuali.

Per tutte le attività, per quanto riguarda il codice Python, è possibile (e gradito) produrre e consegnare un notebook jupyter .ipynb (<https://jupyter.org/>) invece di codice .py e relativi commenti separati su PDF (per comodità di consultazione, consegnare comunque anche una stampa PDF del notebook oltre al notebook stesso).

Scadenza per premio partecipazione: **11 novembre 2022**