

Progetto di Basi di Dati

a.a. 2019/2020

Prof. Vincenzo Moscato

Dato il DATASET, costituito dall'insieme dei file in formato CSV scaricabili da <https://github.com/pcm-dpc/COVID-19/tree/master/dati-province>, e relativo all'andamento del contagio del virus COVID-19 nelle province Italiane, si effettui attraverso l'ambiente ORACLE LIVE SQL (o mediante la distribuzione Oracle XE 18c):

- 1) La creazione di una tabella master (con lo stesso schema dei file in formato CSV) ed il relativo popolamento con i dati del contagio per tutte le province italiane dal 25/02/2020 al 03/05/2020 (utilizzare una qualsiasi strumento di conversione da CSV a SQL come ad esempio il tool <https://www.convertcsv.com/csv-to-sql.htm>);
- 2) La verifica della 3NF per lo schema della tabella precedentemente istanziata e l'eventuale decomposizione con la definizione di tutti i vincoli (mediante comandi DDL);
- 3) L'arricchimento dello schema (mediante comandi di ALTER TABLE) ottenuto con ulteriori informazioni utili all'analisi del fenomeno, ad esempio:
 - a. per ogni data il numero di morti, il numero di ricoveri in strutture ospedaliere, il numero di ricoveri in terapia intensiva, etc.
 - b. per ogni regione il numero di abitanti, la densità abitativa, superficie in km2, numero di autostrade e strade statali, numero di aeroporti e stazioni, etc.;
 - c. per ogni provincia il numero il numero di abitanti, la densità abitativa, il numero di scuole, alberghi/strutture recettive, il numero di ospedali, il numero di spostamenti intra e extra provincia, etc.
 - d. etc.
- 4) L'individuazione, attraverso un processo di reverse engineering, di un possibile schema concettuale E/R della base di dati;
- 5) La specifica in SQL di una serie di query utili all'analisi dell'andamento del contagio del COVID-19 (e.g., numero di contagi per provincia in una determinata finestra temporale, regione col il maggior numero di contagi per densità abitativa, etc.), con un'eventuale visualizzazione grafica dei risultati (es. attraverso l'utilizzo di MS Excel);
- 6) La specifica in PL/SQL di una serie di procedura/trigger che possano consentire un'analisi più flessibile ed eventuali aggiornamenti automatici della base di dati qualora vogliano essere importati dati successivi al 03/5/2020, ad esempio:
 - a. STORED PROCEDURE aventi come parametri di ingresso le differenti variabili dell'analisi (es. giorno o finestra temporale di analisi, provincia o regione da analizzare, etc.);
 - b. TRIGGER che all'atto di inserimenti nella tabella master esegue in automatico il popolamento delle tabelle dello schema normalizzato;
- 7) La definizione di viste sui dati ed indici che possono essere utili per migliorare i tempi di esecuzione delle query.

N.B.: Il progetto può essere svolto da gruppi di al massimo 2 persone e verrà discusso il giorno dell'esame. Un documento pdf contenente la documentazione del progetto il dovrà essere inviato via email entro 2gg prima della seduta di esame.