

## **INTRODUZIONE**

Raccolta documentazione inerente al progetto di Basi di Dati, affidatoci dal Professore Elio Masciari, che prevede la raccolta e l'analisi di informazioni relative all'andamento del virus nCoV-19 nel periodo 25/02/2020-03/05/2020. Il progetto ha visto la partecipazione degli studenti Matteo Caccavale, Manuel Sommella, Domenico Castiello, Luca Pisani. Relazione a cura degli studenti Manuel Sommella e Luca Pisani.

Programmi utilizzati: DataGrip per la gestione del DataBase, MS Excel per la visualizzazione grafica delle query, MS Word per la stesura della relazione, MS Teams per le riunioni a distanza, PasteBin per la corretta visualizzazione e organizzazione dei codici SQL, Microsoft OneDrive & SharePoint per la condivisione del documento Excel raffigurante le query in formato grafico.

Le riunioni sono state fissate ogni Lunedì, Venerdì e Sabato a partire dal 04/05/2020.

### Fonti:

https://github.com/pcm-dpc/COVID-19 (DataSet)

<u>https://www.tuttitalia.it/</u> (Informazioni relative alle scuole e ospedali)

<u>https://www.wikipedia.org/</u> (Informazioni relative alle informazioni GeoPolitiche delle singole province e regioni, dati relativi alle infrastrutture)

http://www.abitareeanziani.it/wp-content/uploads/2016/07/Auser-ricerca-case-riposo.pdf (Informazioni relative alle case di riposo)

## I. CREAZIONE E POPOLAMENTO DELLA MASTER TABLE

La creazione della master table è un processo fondamentale ai fini della comprensione di quali saranno gli attributi chiave della relazione. I passaggi effettuati sono stati i seguenti:

- Comprendere in che formato sono disponibili i dati.
- Determinare in quale modo è possibile inserire i dati all'interno della tabella master.
- Creazione della tabella master attraverso comandi DDL (<a href="https://pastebin.com/8ayjhL6M">https://pastebin.com/8ayjhL6M</a>).
- Utilizzo di tool on-line che consentono la conversione da formato CSV a SQL-INSERT (https://www.convertcsv.com/csv-to-sql.htm).
- Popolamento della tabella attraverso l'utilizzo di DataGrip.

Durante tale processo si sono riscontrate varie anomalie.

La prima è stata riscontrata nella chiave primaria della relazione, prontamente risolta ponendo Data e Codice\_provincia come tali. La ragione di tale scelta è stata dettata dal fatto che sia Data, che Codice\_provincia, pur essendo attributi primari della tabella, non rispettavano i vincoli di chiave singolarmente, ripetendosi più volte.

Un'altra anomalia si è presentata riguardo l'attributo long che, essendo un comando SQL, si è visto ri-denominare in longitudine.

Si è infine deciso di eliminare gli attributi Stato, Note\_it e Note\_en, non ritenuti necessari ai fini del progetto.

Analisi andamento del contagio del virus nCoV-19

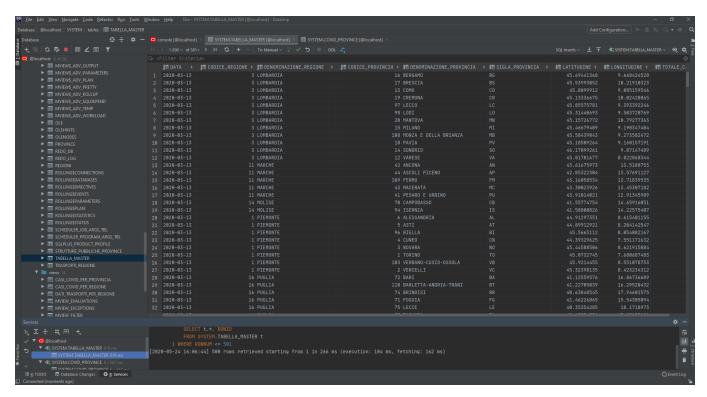


Tabella Master

# II. NORMALIZZAZIONE DELLA MASTER TABLE RISPETTANDO LA 3NF

Come da titolo, la normalizzazione è stata effettuata rispettando la 3NF. I passaggi effettuati sono:

- Analisi della tabella master atta all'identificazione degli attributi chiave.
- Identificazione delle dipendenze funzionali.
- Normalizzazione dapprima in 2NF e successivamente in 3NF.
- Creazione delle relative tabelle in linguaggio DDL (<a href="https://pastebin.com/yURRFZAy">https://pastebin.com/yURRFZAy</a>).
- Popolamento delle tabelle attraverso comandi SQL-INSERT ottenuti attraverso il convertitore del precedente punto.

La normalizzazione ha portato con sé alcuni problemi, che riguardavano soprattutto il popolamento.

La prima anomalia si è presentata riguardava il codice\_regione, notando come P.A. Bolzano e P.A. Trento condividessero lo stesso codice, nonostante non appartenessero alla stessa regione, risultando in un'infrazione del vincolo di chiave primaria.

Tale problema è stato risolto effettuando una somma dei campi Totale\_Casi, accorpando tali province autonome in una sola regione, Trentino-Alto Adige, con codice\_regione = 4.

Un altro problema ha riguardato il popolamento delle tabelle in sé: lo scopo era quello di riempire le tabelle senza eliminare a mano i valori di attributi non necessari.

Si è poi scoperto come il convertitore fornisse una soluzione, eliminando gli attributi e i conseguenti valori non necessari attraverso delle opzioni di output. In questo modo la popolazione delle tabelle è stata relativamente semplice e molto rapida.

Data della relazione: 23/05/2020

Analisi andamento del contagio del virus nCoV-19

## Dipendenze funzionali tabella master:

 Data, codice\_provincia → Totale\_casi, note\_it, note\_en, Denominazione\_Provincia, Sigla\_Provincia, lat, long, Denominazione\_Regione, Stato

## Dipendenze funzionali tabella in 2NF:

- Data, provincia → Totale\_casi, note\_it, note\_en
- Codice\_provincia → Denominazione\_Provincia, Sigla\_Provincia, lat, long, Codice\_regione, Denominazione\_Regione, Stato

## Dipendenze funzionali tabella in 3NF:

- Data, provincia → Totale\_casi, note\_it, note\_en
- Codice\_provincia → Denominazione\_Provincia, Sigla\_Provincia, lat, long
- Codice\_regione → Denominazione\_Regione, Stato

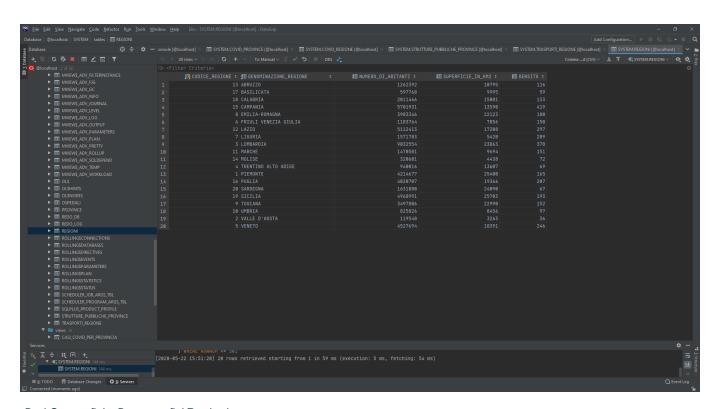
Analisi andamento del contagio del virus nCoV-19

# III. ARRICCHIMENTO DELLO SCHEMA

All'interno delle relazioni era possibile includere numerosi altri attributi che potessero arricchire le informazioni relative al contagio. Tali modifiche sono state effettuate mediante comandi di ALTER TABLE, e ciò ha potuto permettere di aggiungere, ad esempio, il numero di casi per regione, numero di ospedalizzati, informazioni relative alla geografia e demografia della regione e della provincia. Le tabelle possono essere consultate a questo indirizzo:

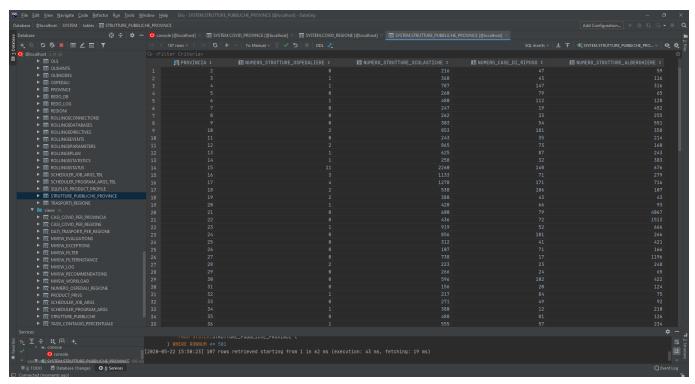
https://pastebin.com/aV1qXLy6

Alcuni screenshot delle tabelle visualizzate su DataGrip:

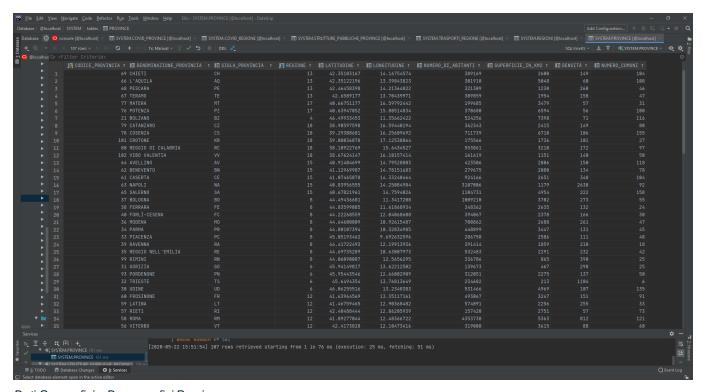


Dati Geografici e Demografici Regioni

Analisi andamento del contagio del virus nCoV-19



Strutture Pubbliche per Provincia

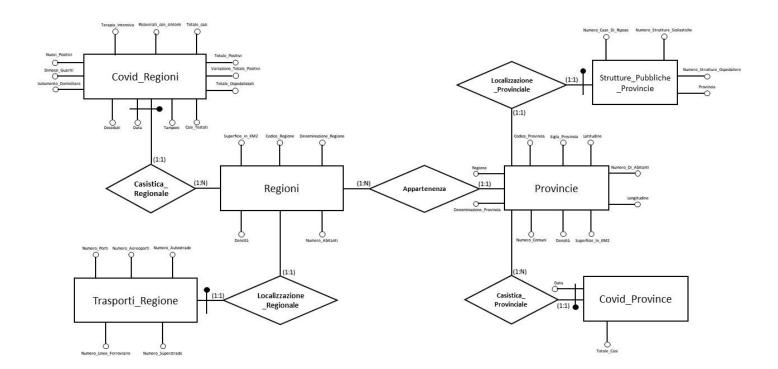


Dati Geografici e Demografici Province

# IV. SCHEMA CONCETTUALE E/R

Attraverso un processo di reverse engineering è stato possibile scrivere lo schema relazionale e lo schema concettuale E/R della base di dati:

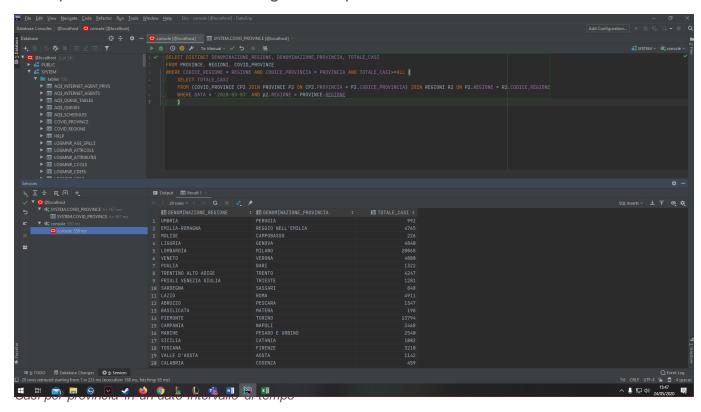
## Schema E/R



Analisi andamento del contagio del virus nCoV-19

# V. QUERY E VISUALIZZAZIONE GRAFICA CON MS EXCEL

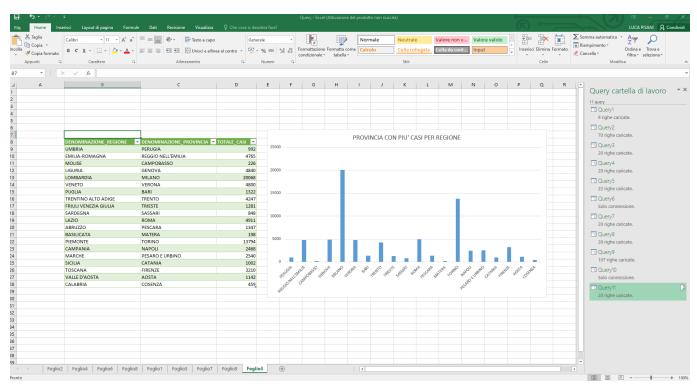
E' stato necessario sviluppare una serie di query di fondamentale importanza al fine di effettuare un'analisi approfondita dei dati in relazione all'andamento del contagio del nCoV-19. A seguito dello sviluppo è stato possibile, attraverso la funzione "Dati → Nuova Query" di Microsoft Excel, l'importazione in foglio di calcolo delle tabelle selezionate dalle query e la relativa rielaborazione dei dati in esse presente al fine di creare grafici esplicativi.



E' possibile trovare altre query a questo indirizzo: <a href="https://pastebin.com/xw5nReHp">https://pastebin.com/xw5nReHp</a>

# IMPLEMENTAZIONE DI MS EXCEL

Di seguito è riportata un'immagine della relativa rielaborazione in foglio di calcolo MS Excel della query:



Casi per provincia in un dato intervallo di tempo

E' possibile trovare il documento Excel (esportato in foglio di calcolo google) con l'elaborazione grafica delle altre query a questo indirizzo:

# https://communitystudentiunina-

<u>my.sharepoint.com/:x:/g/personal/luca\_pisani\_studenti\_unina\_it/Ec0hzTS2JKIInlmhfM4zwDcBsDxa0DiSA9d4gAMaW7TLjA?e=6F026v</u>

Analisi andamento del contagio del virus nCoV-19

# VI. PL/SQL: PROCEDURE E TRIGGER

Al fine di consentire un'analisi più flessibile della base di dati sono stati implementate alcune procedure e trigger:

- Trigger → <a href="https://pastebin.com/LMgJRKs7">https://pastebin.com/LMgJRKs7</a>
- Procedure → <a href="https://pastebin.com/YiPKGPsN">https://pastebin.com/YiPKGPsN</a>

Un esempio di chiamata:

```
| The part yee Brown Control Control | Decorate | Decor
```

Numero di contagi per provincia in una data finestra temporale

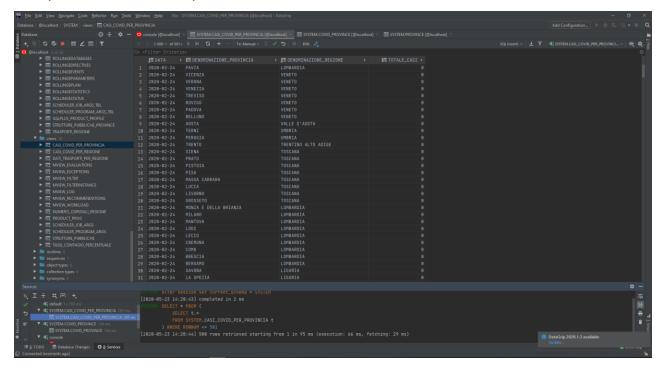
Analisi andamento del contagio del virus nCoV-19

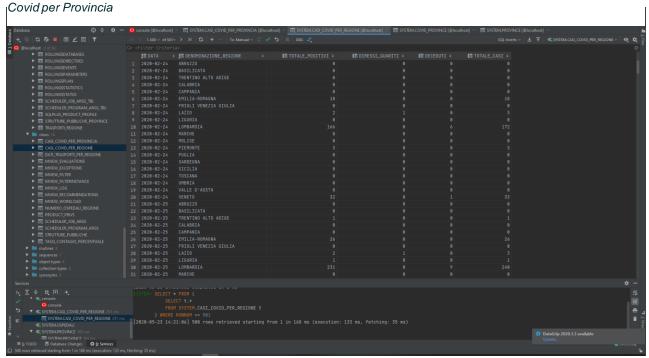
# **VII. VISTE**

Sono state definite e sviluppate delle viste sui dati, utili al miglioramento del tempo di esecuzione delle query. E' possibile trovare le viste qui:

# https://pastebin.com/g0LeaSyn

Se ne riportano alcuni esempi:





Covid per Regione