



Universidade do Minho
Escola de Engenharia

Análise de Dados

TP1

Gonçalo Rodrigues Pinto - a83732

Repositório da situação pandémica na EU/EEA

- Este exercício prático tem como objetivo desenvolver uma possível base de dados para um repositório COVID- 19 da EU/EEA.

Ferramentas: MySQL e MySQL Workbench

Requisitos a desenvolver

1. Criar um esquema conceptual para a base de dados do repositório:
 - Com pelo menos duas entidades;
 - Com os seus respetivos atributos e relacionamentos;
2. Converter o modelo conceptual para o seu modelo lógico;
3. Converter o modelo lógico para o seu modelo físico;
4. Implementar pelo menos um *trigger* e uma função;
5. Povoar as tabelas com os dados presentes nos ficheiros covid19_eu_eea_uk_data.csv e covid19_eu_eea_uk_testing.csv;
6. Criar pelo menos 3 *queries* SQL cujo resultado seja um indicador interpretativo sobre o tema;
7. Desenhar e detalhar o processo desenvolvido para chegar à solução pretendida.

1. Esquema Conceptual

Entidades:

-> *Country*

Atributos:

- *country_code* (chave primária)
- *name*
- *countryterritoryCode*
- *population*

-> *Week*

- *idWeek* (chave primária)

Relacionamentos:

Country com *Week* (de Muitos para Muitos) -> *Testing*

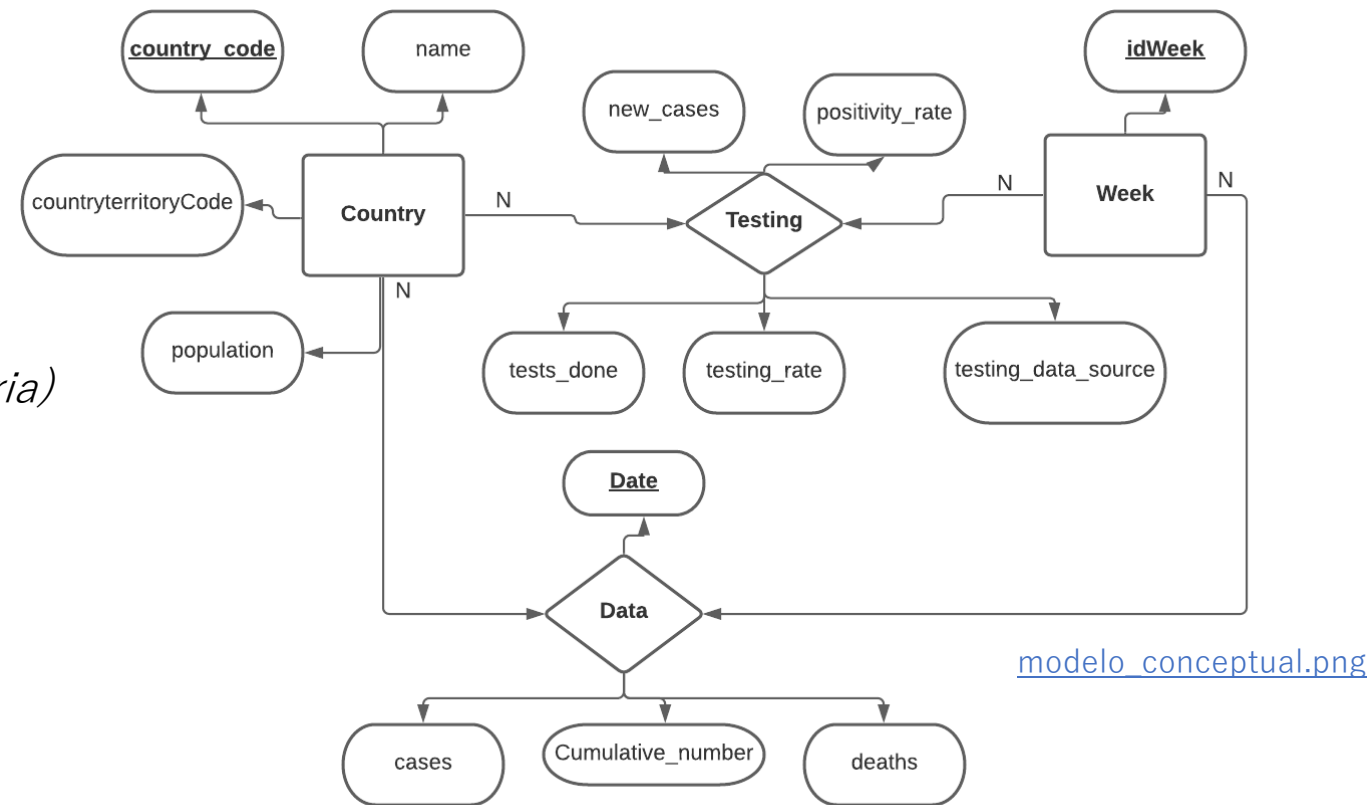
Country com *Week* (de Muitos para Muitos) -> *Data*

Nome:

Atributos dos Relacionamentos:

- *new_cases* ; - *tests_done* ;
- *positivity_rate* ; - *testing_rate* ;
- *testing_data_source* ;

- *cases* ; - *deaths* ;
- *Cumulative_number* ;

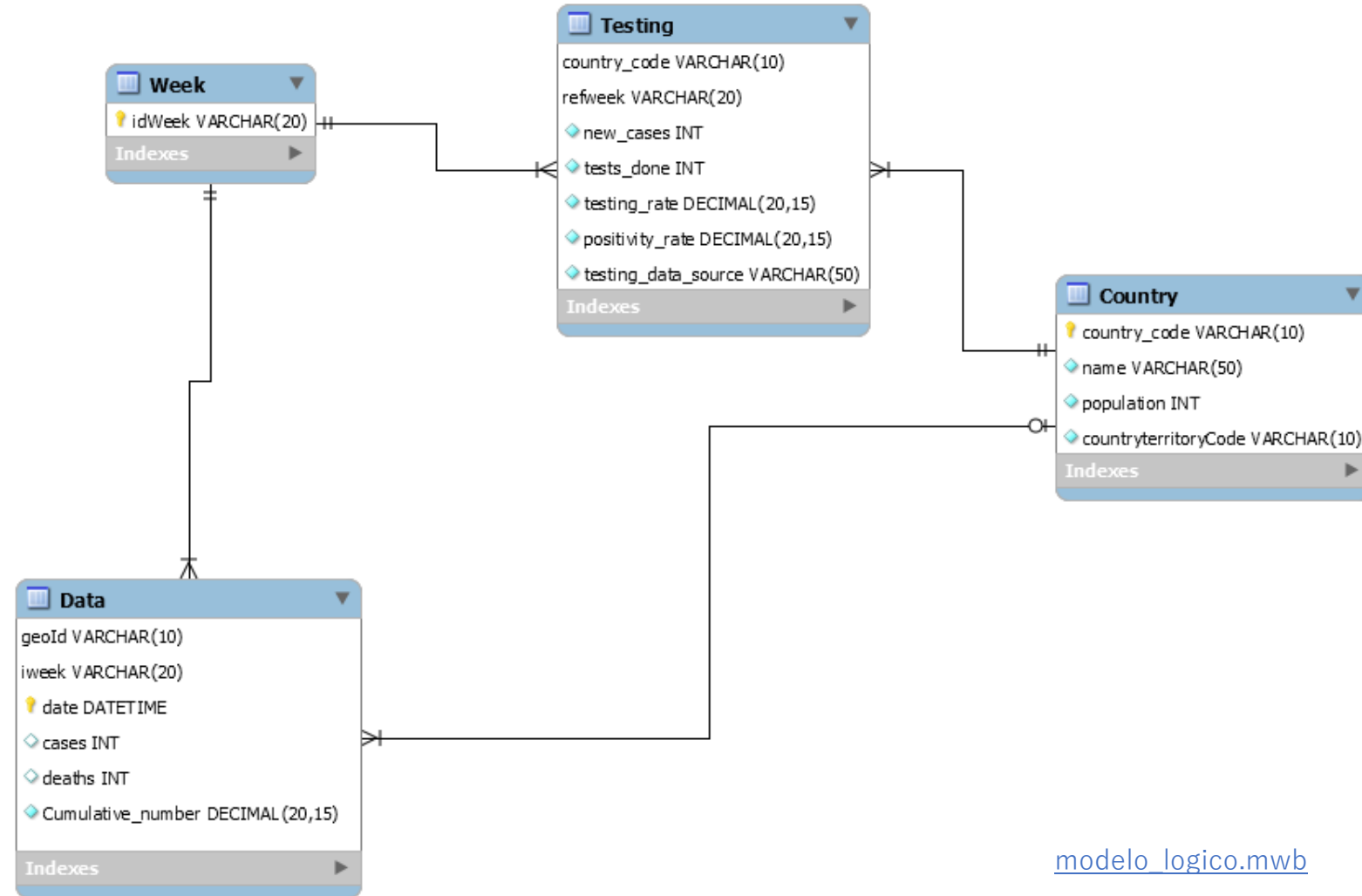


1. Esquema Conceptual - Justificação

Considerando os ficheiros covid19_eu_eea_uk_data.csv e covid19_eu_eea_uk_testing.csv pude constatar que a informação que se encontrava duplicada era a referência ao país como também a semana. Portanto decidi criar como entidades estas duas informações no caso do país os atributos deste considereei a informação referida nas tabelas no caso da semana apenas o número desta. Tal como foi referido estas entidades relacionam através dos ficheiros apresentados por isso cada relacionamento corresponde ao ficheiro em causa que possui como atributos do relacionamento os atributos que estão registados nos csv's.

2. Modelo conceptual para o seu modelo lógico

Partindo do modelo conceptual criado e utilizando MySQL Workbench criei o modelo lógico apresentado onde os relacionamentos entre as entidades previamente referidas. Como os relacionamentos entre estas eram de muitos para muitos gerou-se tabelas onde a informação desta corresponde aos atributos dos relacionamentos respetivos.



[modelo_logico.mwb](#)

3. Modelo lógico para o seu modelo físico

Partindo do modelo criado no MySQL Workbench como foi dito anteriormente, para traduzir o esquema lógico para o esquema físico automaticamente utilizou-se a opção “forward engineer”, como pode ser observado no ficheiro modelo_físico.sql, que levou a criação do Sistema de Gestão de Banco de Dados .

[modelo_fisico.mwb](#)

4. Implementação de um *trigger* e de uma função

Função implementada - A partir de um nome de um país saber o número total de mortes registados:

```
DELIMITER $$  
CREATE FUNCTION tot_Deaths(country_name varchar(50)) Returns int  
Deterministic  
BEGIN  
    Declare Total_Mortes int;  
    select sum(d.deaths) into Total_Mortes from data d, country c where c.name=country_name and d.geold=c.country_code;  
    RETURN Total_Mortes;  
END $$
```

Trigger implementado - Atualizar a populacao de um país após insercao de mortes:

```
DELIMITER $$  
create trigger updatePopulation after insert on data for each row  
begin  
    Update country c set NEW.c.population=OLD.c.population-NEW.data.deaths where c.country_code = d.geold ;  
end $$
```

5. Povoamento as tabelas com os dados presentes nos ficheiros

De forma a povoar as tabelas com os dados presentes nos ficheiros `covid19_eu_eea_uk_data.csv` e `covid19_eu_eea_uk_testing.csv`, criou-se um programa em **Java** disponível que permitiu importar os dados relativos ao covid19 da União Europeia. Este sistema efetuou a leitura dos diferentes *datasets* referentes aos dados e aos seus testes, processou a informação e posteriormente escreveu num ficheiro que representa o script para povoamento.

[povoamento.sql](#)

6. Queries SQL cujo resultado são um indicador interpretativo sobre o tema

#Querie1 - Saber quantos testes os data_source fizeram:

```
select t.testing_data_source as Data_Source, count(1) as Tests from testing t group by Data_Source order by Tests desc ;
```

#Querie2 - Determinar a média de novos casos e testes feitos por país:

```
Select c.name as Country, c.population as Population, avg(t.new_cases) as AVG_NewCases, avg(t.tests_done) as AVG_TestsDone from country c, testing t where c.country_code=t.country_code group by c.name;
```

#Querie3 - Mostrar os 3 países que mais casos tiveram:

```
select c.countryterritoryCode as CTC, c.name as Country, c.population as Population, sum(d.cases) as Tot_Cases from country c, data d where c.country_code=d.geold group by Country order by Tot_Cases desc limit 3;
```