



Universidade do Minho
Escola de Engenharia

Universidade do Minho
Mestrado Integrado em Engenharia Informática

Perfil EC

Análise de Dados

TP 2

Outubro 2020

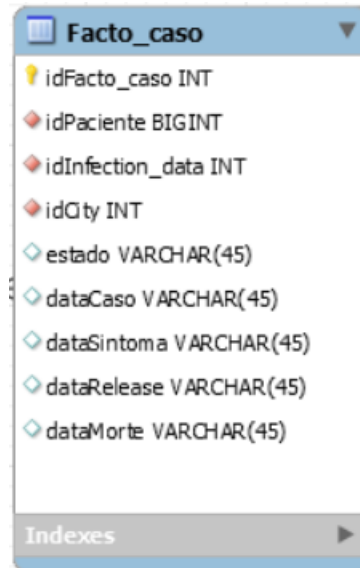
Tânia Filipa Amorim Rocha
A85176

1 Introdução

Este relatório foi desenvolvido no âmbito do primeiro guião prático da UC análise de dados. Será discutido e apresentado detalhadamente o processo de construção de um modelo em estrela e implementações extra com base em dados fornecidos em ficheiros csv sob o contexto das infeções por covid-19 na Coreia do Sul.

2 Análise e Construção

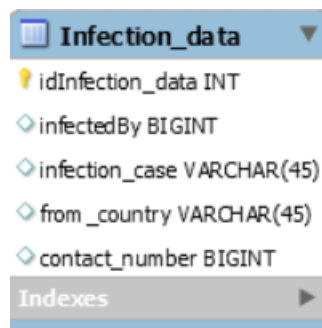
Tendo em conta os dados descritos no csv, foi optado por ter uma tabela de factos que retrata as informações gerais do modelo, isto é, contém um id gerado para o caso em questão e tem informações sobre o paciente, datas pertinentes e informações adicionais.



Facto_caso	
idFacto_caso	INT
idPaciente	BIGINT
idInfection_data	INT
idCity	INT
estado	VARCHAR(45)
dataCaso	VARCHAR(45)
dataSintoma	VARCHAR(45)
dataRelease	VARCHAR(45)
dataMorte	VARCHAR(45)

Figura 1: Tabela de Factos

Para além da tabela de factos, foi desenvolvida 3 tabelas de dimensão adicionais, nomeadamente, uma tabela Paciente, que contém informação sobre o mesmo como por exemplo idade, sexo e o seu número de identificação, uma tabela com informação quanto à infecção por covid que contém informação sobre onde e porque foi contaminado e em que contexto essa infecção ocorreu. Por fim uma tabela de dimensão referente à cidade de onde o paciente pertence incluindo província.



Infection_data	
idInfection_data	INT
infectedBy	BIGINT
infection_case	VARCHAR(45)
from_country	VARCHAR(45)
contact_number	BIGINT

Figura 2: Tabela de dimensão das informações de infecção

3 Modelo em Estrela

O resultado final do modelo é o representado a seguir:

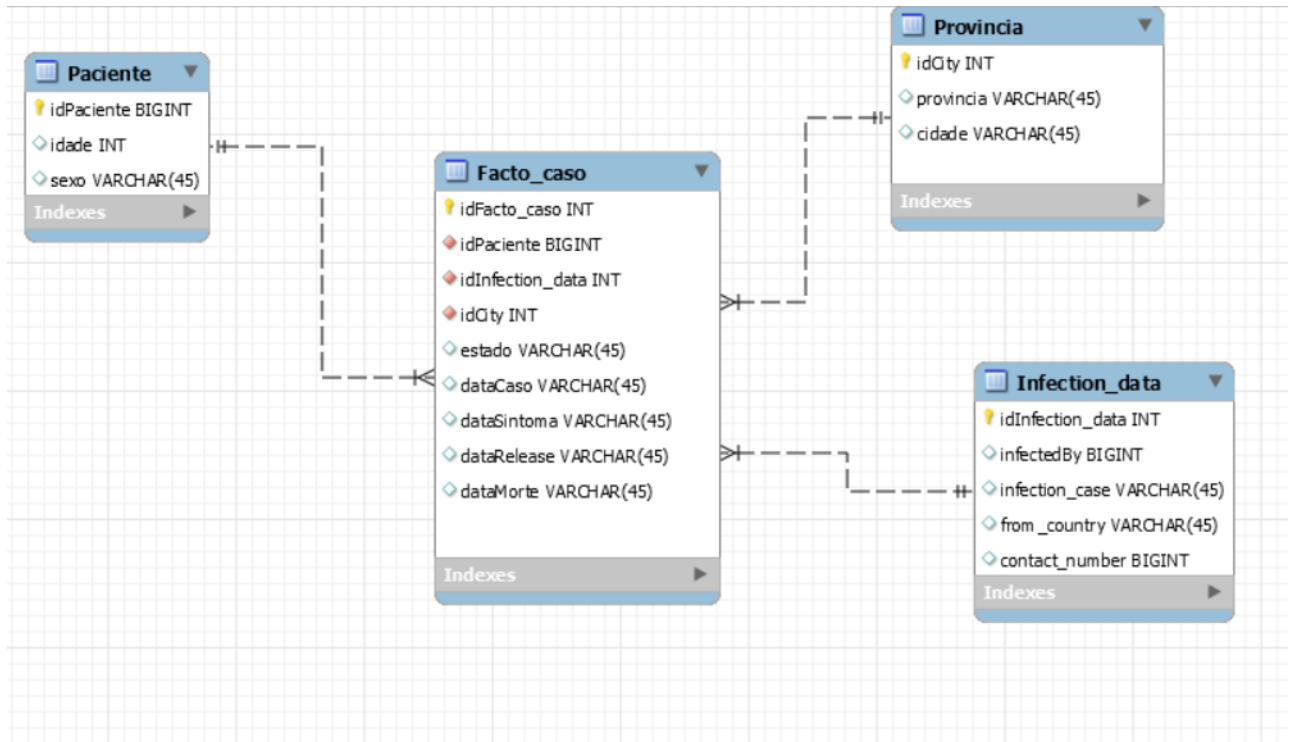


Figura 3: Modelo em estrela

4 Queries e Teste

Na medida a testar a base de dados construída e posteriormente povoada foram desenvolvidas 3 queries:

- Query que lista os pacientes por número de contactos

```
select t.idPaciente, a.contact_number as Contactos from facto_caso t, infection_data a
where t.idInfection_data = a.idInfection_data
group by t.idPaciente
order by Contactos desc;
```

	idPaciente	Contactos
▶	6100000098	6100000099
	6100000097	6100000098
	1000000819	1000000796
	1200000031	1160
	1100000001	1091
	1100000007	485
	1300000001	450
	2000000003	422
	1100000012	296
	2000000006	290
	1100000128	185
	4100000000	174

Figura 4: query 1 e respectivo resultado

- Query que lista os pacientes que faleceram

```
2 • select t.idPaciente , t.dataMorte from facto_caso t where t.estado='deceased';
```

Figura 5: Query 2

	idPaciente	dataMorte
	1200000038	23/02/20
	1200000114	26/02/20
	1200000620	02/03/20
	1200000901	04/03/20
	1200001064	01/03/20
	1200001288	02/03/20
	1200001443	27/02/20
	1200001921	05/03/20
	1200002085	04/03/20
	1200002289	05/03/20
	1200002614	01/03/20
	1200002769	02/03/20

Figura 6: Resultado da querie 2

- Querie que lista todos os pacientes e as suas províncias de uma determinada cidade, neste caso Busan.

```
6 • select t.idPaciente , a.provincia from facto_caso t, provincia a where t.idCity=a.idCity and a.cidade='Busan';
```

Figura 7: Querie 3

	idPaciente	provincia
▶	1100000001	Dongnae-gu
	1100000002	Haeundae-gu
	1100000003	Haeundae-gu
	1100000004	Haeundae-gu
	1100000005	Haeundae-gu
	1100000006	Yeonje-gu
	1100000007	
	1100000008	Dongnae-gu
	1100000009	Dongnae-gu
	1100000010	Dongnae-gu
	1100000011	Buk-gu
	1100000012	Nam-gu

Figura 8: Resultado da querie 3

5 Conclusão

É assim concluído este relatório de descrição de construção de um modelo em estrela do mysql com dados fornecidos em formato csv e respectivas queries para teste.