

Disciplina – Banco de Dados Não Relacional - DSM

Primeira Atividade Avaliativa

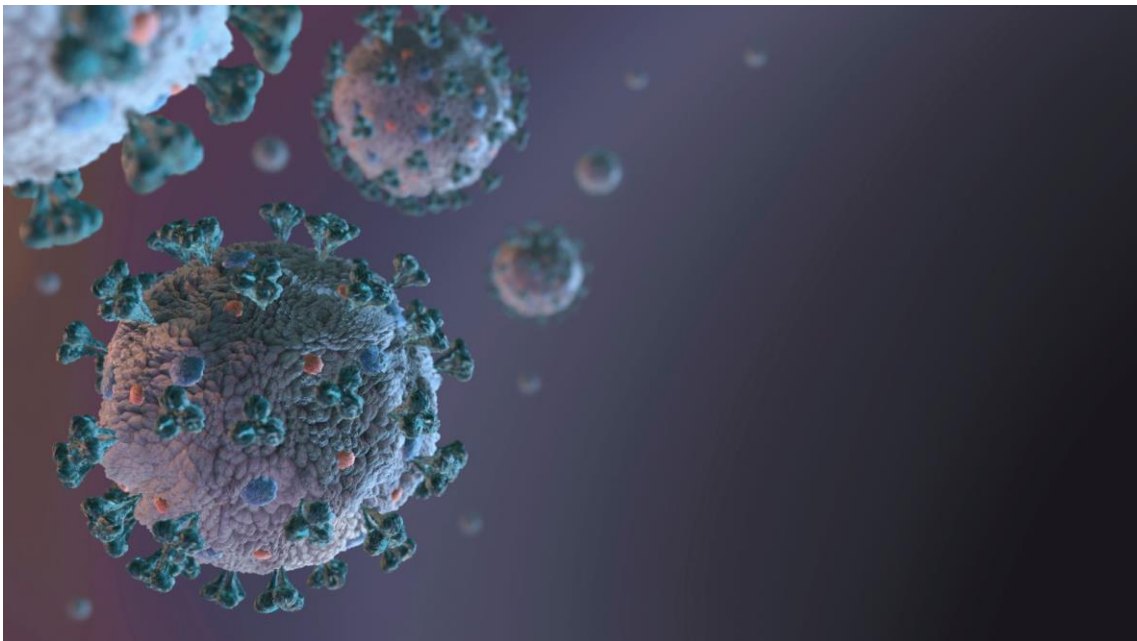
Marilia P Nascimento de Godoy

INTRODUÇÃO

Em 11 de março de 2020, a COVID-19 foi caracterizada pela OMS como uma pandemia. O termo “pandemia” se refere à distribuição geográfica de uma doença e não à sua gravidade. A designação reconhece que existiu surtos de COVID-19 em vários países e regiões do mundo.

A criação desse banco tem por objetivo disseminar dados sobre a pandemia de Covid-19 em forma de fácil compreensão e que possibilite a comparação de tendências. Permitido o acompanhamento da evolução da pandemia no mundo.

As diferentes metodologias e fontes de dados utilizadas podem apresentar diferentes resultados. A avaliação conjunta e a busca de consensos são, portanto, indispensáveis.



Descrição do processo de elaboração DATA WAREHOUSE

Para uma modelagem multidimensional de data warehouse para dados da COVID-19 em um banco de dados não relacional, você pode considerar os seguintes pontos:

Identificação das Dimensões:

Identifiquei as dimensões relevantes para sua análise. Isso pode incluir dimensões como data, localização geográfica (país, estado, cidade), tipo de caso (confirmado, recuperado, fatal), faixa etária, sexo, tipo de teste, etc.

Fatos e Métricas:

Identifiquei as métricas desejadas para analisar. Isso inclui o número de casos confirmados, número de mortes, número de recuperados, taxa de positividade dos testes, etc. Estas serão suas medidas ou fatos.

Projeto da Tabela de Fatos:

Criei uma tabela de fatos que conterá as métricas medidas em relação às dimensões. Cada linha na tabela de fatos representa uma combinação de valores das dimensões em um determinado ponto no tempo.

Projeto das Tabelas de Dimensão:

Criei tabelas de dimensão para cada uma das dimensões identificadas. Cada tabela de dimensão deve conter informações detalhadas sobre essa dimensão.

Modelagem Hierárquica:

Considere as hierarquias em suas dimensões, por exemplo, a hierarquia temporal ou hierarquia geográfica (país, estado, cidade, etc.).

Carga e Atualização de Dados:

Objetivo desenvolver um processo para carregar e atualizar os dados no data warehouse. Isso pode envolver a extração de dados de fontes externas (como repositórios de dados públicos), transformação dos dados conforme necessário e carga no banco de dados não relacional.

Consulta e Análise:

Implementei consultas para extrair informações úteis do data warehouse. Isso pode envolver a criação de consultas SQL ou a utilização de ferramentas de visualização de dados.

Ferramentas Utilizadas

Utilizei o MySQL Workbench é uma ferramenta visual muito popular para o design, desenvolvimento e administração de bancos de dados MySQL. E editor SQL integrado que ajuda na escrita, execução e depuração de queries SQL.

As minhas tabelas dimensões foram adicionados de arquivos csv retiradas do site :

<https://www.kaggle.com/>

Tabelas

As tabelas dimensões foram criada no MySQL Workbench através da importação de arquivos csv. Realizadas somente as alterações de chave primarias e tipos das variáveis. Ficando assim:

```
ALTER TABLE `covid19`.`data`  
  
CHANGE COLUMN `Id_data` `Id_data` INT NOT NULL AUTO_INCREMENT ,  
  
CHANGE COLUMN `Data` `Data` DATETIME NOT NULL ,  
  
ADD PRIMARY KEY (`Id_data`);  
  
;
```

```
ALTER TABLE `covid19`.`dim_regiaooms`  
  
CHANGE COLUMN `id_regiao` `id_regiao` INT NOT NULL AUTO_INCREMENT ,  
  
CHANGE COLUMN `Regiao_OMS` `Regiao_OMS` VARCHAR(100) NOT NULL ,  
  
ADD PRIMARY KEY (`id_regiao`);  
  
, RENAME TO `covid19`.`dim_regiao_oms` ;
```

```
ALTER TABLE `covid19`.`dim_medidor_mundial`  
  
ADD COLUMN `dim_medidor_mundialcol` VARCHAR(45) NOT NULL AFTER `Regiao_ONU`,  
  
CHANGE COLUMN `Id_med` `Id_med` INT NOT NULL AUTO_INCREMENT ,  
  
CHANGE COLUMN `Pais_regiao` `Pais_regiao` VARCHAR(100) NULL DEFAULT NULL ,  
  
ADD PRIMARY KEY (`Id_med`);  
  
;
```

```

ALTER TABLE `covid19`.`dim_covid_19_completa`
ADD COLUMN `dim_covid_19_completacol` VARCHAR(45) NOT NULL AFTER `Regiao_OMS`,
CHANGE COLUMN `Id_com` `Id_com` INT NOT NULL AUTO_INCREMENT ,
CHANGE COLUMN `Pais_regiao` `Pais_regiao` VARCHAR(100) NOT NULL ,
CHANGE COLUMN `Date` `Date` DATETIME NOT NULL ,
CHANGE COLUMN `Regiao_OMS` `Regiao_OMS` VARCHAR(100) NULL DEFAULT NULL ,
ADD PRIMARY KEY (`Id_com`);
;

```

```

ALTER TABLE `covid19`.`dim_covid_19_completa`
ADD COLUMN `dim_covid_19_completacol` VARCHAR(45) NOT NULL AFTER `Regiao_OMS`,
CHANGE COLUMN `Id_com` `Id_com` INT NOT NULL AUTO_INCREMENT ,
CHANGE COLUMN `Pais_regiao` `Pais_regiao` VARCHAR(100) NOT NULL ,
CHANGE COLUMN `Date` `Date` DATETIME NOT NULL ,
CHANGE COLUMN `Regiao_OMS` `Regiao_OMS` VARCHAR(100) NULL DEFAULT NULL ,
ADD PRIMARY KEY (`Id_com`);
;

```

Já a tabela fato, criei através de script SQL no Workbench:

```

CREATE TABLE Fato_Covid19 (
    Fato_Id INT AUTO_INCREMENT PRIMARY KEY,
    Id_data INT,
    Id_com INT,
    Id_med INT,
    id_regiao INT,

```

```

Confirmado BIGINT,
Mortes BIGINT,
Recuperados BIGINT,
Ativos BIGINT,
Total_Casos BIGINT,
Total_mortos BIGINT,
Total_Recuperados BIGINT,
Casos_Ativo BIGINT,
Serios_Criticos BIGINT,
Total_Testes BIGINT,
FOREIGN KEY (Id_data) REFERENCES dim_data(Id_data),
FOREIGN KEY (Id_com) REFERENCES dim_covid_19_completa(Id_com),
FOREIGN KEY (Id_med) REFERENCES dim_medidor_mundial(Id_med),
FOREIGN KEY (id_regiao) REFERENCES dim_regiao_oms(id_regiao)
);

```

Script em anexo.

Para inserir os dados na tabela fato, utilizei o script, também em anexo:

```

INSERT INTO Fato_Covid19 (
    Id_data, Id_com, Id_med, id_regiao, Confirmado, Mortes, Recuperados, Ativos,
    Total_Casos, Total_mortos, Total_Recuperados, Casos_Ativo, Serios_Criticos, Total_Testes
)
SELECT
    d.Id_data, c.Id_com, m.Id_med, r.id_regiao, c.Confirmado, c.Mortes, c.Recuperados,
    c.Ativos,
    m.Total_Casos, m.Total_mortos, m.Total_Recuperados, m.Casos_Ativo, m.Serios_Criticos,
    m.Total_Testes

```

FROM

dim_covid_19_completa c

JOIN

dim_data d ON STR_TO_DATE(c.Date, '%d/%m/%Y') = d.Data

JOIN

dim_medidor_mundial m ON c.Pais_regiao = m.Pais_regiao

JOIN

dim_regiao_oms r ON c.Regiao_OMS = r.Regiao_OMS;

Diagrama da arquitetura (OLTP)

A arquitetura OLTP pode ser descrita em termos de componentes de hardware e software que apoiam o processamento de transações:

- Servidores de Banco de Dados
- Servidores de Aplicação
- Rede
- Clientes
- Esboço de um Diagrama de Arquitetura OLTP

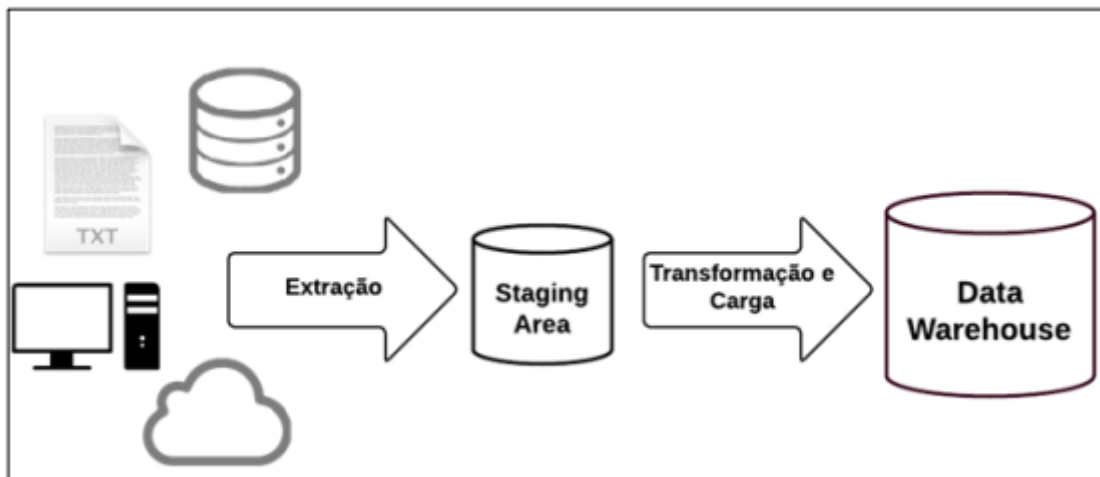
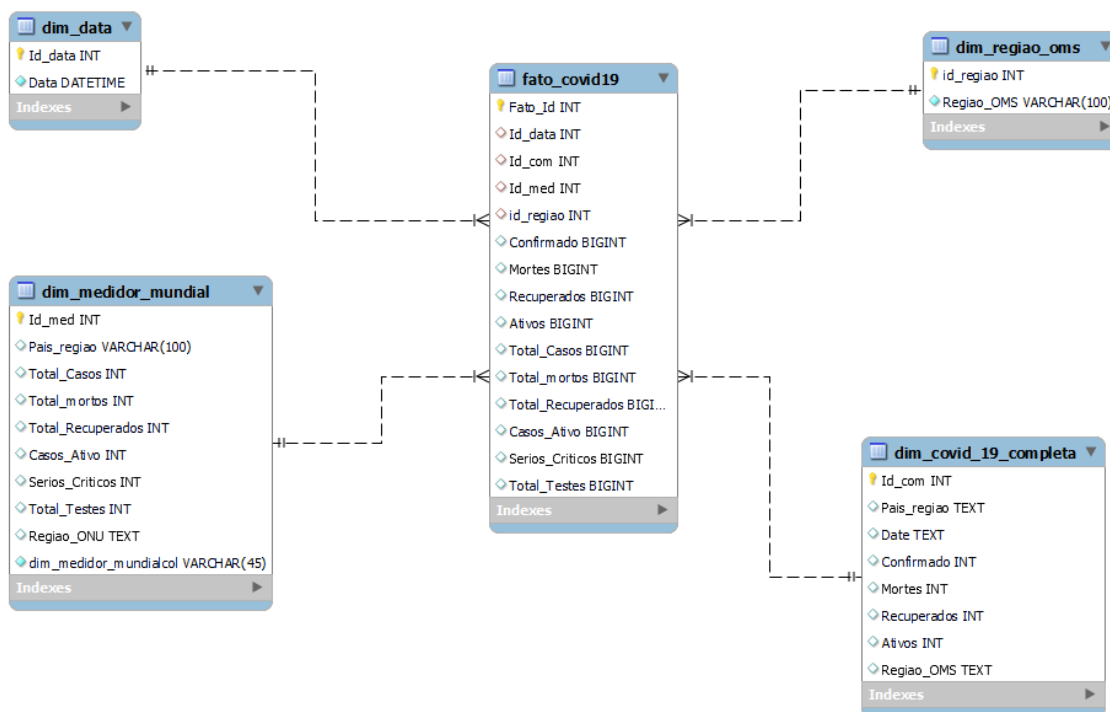


Diagrama multidimensional – Star schema

Criar um diagrama de relacionamento entre suas tabelas de dimensão e a tabela fato é uma ótima maneira de visualizar as interconexões e dependências dentro de seu armazém de dados ou base de dados analítica. Para as tabelas esbocei como seria esse diagrama considerando as relações típicas em um modelo de estrela, comum em esquemas de data warehouse.



Componentes do Diagrama

Tabela Fato (Fato_Covid19): Esta é a tabela central no diagrama de estrela, contendo medidas quantitativas e chaves estrangeiras que apontam para cada uma das tabelas de dimensão.

Tabelas de Dimensão: Incluem **dim_data**, **dim_covid_19_completa**, **dim_medidor_mundial**, e **dim_regiao_oms**.

Descrição das Relações

Fato_Covid19 e dim_data:

Chave Estrangeira: Id_data em Fato_Covid19 que se relaciona com Id_data em dim_data.

Tipo de Relação: Um-para-muitos (1:N) – Muitas entradas na tabela fato podem se relacionar a uma única entrada na tabela de dimensão de data.

Fato_Covid19 e dim_covid_19_completa:

Chave Estrangeira: Id_com em Fato_Covid19 que se relaciona com Id_com em dim_covid_19_completa.

Tipo de Relação: Um-para-muitos (1:N) – Muitas entradas na tabela fato podem se relacionar a uma única entrada na tabela de dimensão de detalhes de COVID-19.

Fato_Covid19 e dim_medidor_mundial:

Chave Estrangeira: Id_med em Fato_Covid19 que se relaciona com Id_med em dim_medidor_mundial.

Tipo de Relação: Um-para-muitos (1:N) – Muitas entradas na tabela fato podem se relacionar a uma única entrada na tabela de dimensão de medição mundial.

Fato_Covid19 e dim_regiao_oms:

Chave Estrangeira: id_regiao em Fato_Covid19 que se relaciona com id_regiao em dim_regiao_oms.

Tipo de Relação: Um-para-muitos (1:N) – Muitas entradas na tabela fato podem se relacionar a uma única entrada na tabela de dimensão de região da OMS.

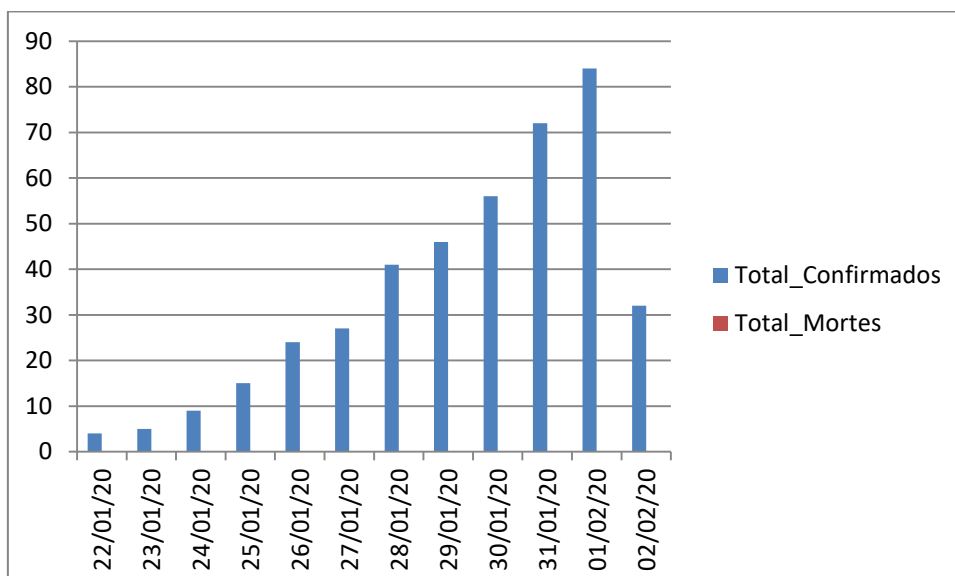
Consultas realizadas:

Tendência Temporal dos Casos e Mortes

Analisar como os casos confirmados, mortes, e recuperações evoluíram ao longo do tempo. Isso pode ajudar autoridades de saúde a entender a dinâmica da pandemia e avaliar a eficácia das medidas de contenção.

```
SELECT d.Data, SUM(f.Confirmado) AS Total_Confirmados, SUM(f.Mortes) AS Total_Mortes  
  
FROM Fato_Covid19 f  
  
JOIN dim_data d ON f.Id_data = d.Id_data  
  
GROUP BY d.Data  
  
ORDER BY d.Data;
```

	Data	Total_Confirmados	Total_Mortes
►	2020-01-22 00:00:00	4	0
	2020-01-23 00:00:00	5	0
	2020-01-24 00:00:00	9	0
	2020-01-25 00:00:00	15	0
	2020-01-26 00:00:00	24	0
	2020-01-27 00:00:00	27	0
	2020-01-28 00:00:00	41	0
	2020-01-29 00:00:00	46	0
	2020-01-30 00:00:00	56	0
	2020-01-31 00:00:00	72	0
	2020-02-01 00:00:00	84	0
	2020-02-02 00:00:00	32	0



Análise de Casos por Região ou País

Comparar o número de casos e taxas de mortalidade entre diferentes regiões ou países. Essa análise pode revelar áreas de alto risco e ajudar na alocação de recursos.

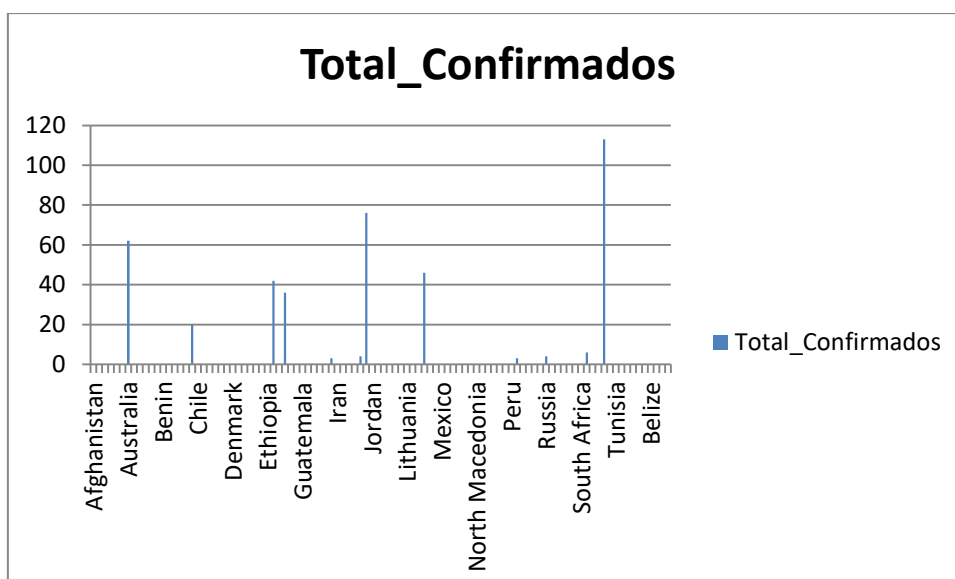
```
SELECT c.Pais_regiao, SUM(f.Confirmado) AS Total_Confirmados, SUM(f.Mortes) AS  
Total_Mortes
```

```
FROM Fato_Covid19 f
```

```
JOIN dim_covid_19_completa c ON f.Id_com = c.Id_com
```

```
GROUP BY c.Pais_regiao;
```

Pais_regiao	Total_Confirmados	Total_Mortes
Germany	36	0
Ghana	0	0
Greece	0	0
Guatemala	0	0
Guinea	0	0
Guyana	0	0
Honduras	0	0
Hungary	0	0
India	3	0
Iran	0	0
Iraq	0	0
Ireland	0	0
Israel	0	0
Italy	4	0
Japan	76	0
Jordan	0	0
Kazakhstan	0	0
Kenya	0	0
Kuwait	0	0
Kyrgyzstan	0	0
Lebanon	0	0
Lithuania	0	0
Luxembourg	0	0
Madagascar	0	0
Malaysia	46	0



Capacidade de Testagem e Sua Relação com Casos Confirmados

Estudar a relação entre o número de testes realizados e os casos confirmados pode ajudar a entender se uma região está testando suficientemente sua população.

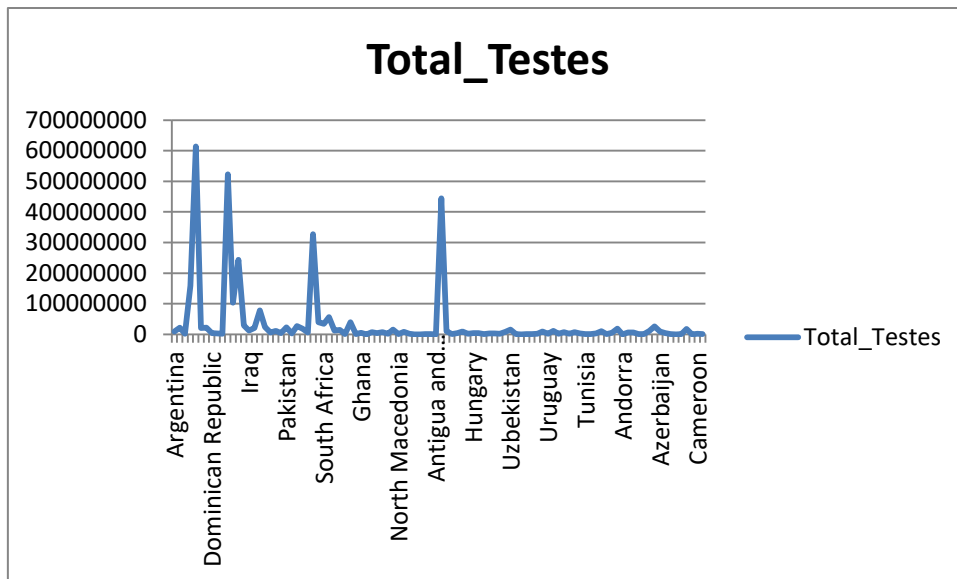
```
SELECT m.Pais_regiao, SUM(m.Total_Testes) AS Total_Testes, SUM(f.Confirmado) AS  
Casos_Confirmados
```

```
FROM Fato_Covid19 f
```

```
JOIN dim_medidor_mundial m ON f.Id_med = m.Id_med
```

```
GROUP BY m.Pais_regiao;
```

Result Grid			
		Filter Rows:	Expo
	Pais_regiao	Total_Testes	Casos_Confirmados
▶	Argentina	9534528	0
	Belgium	21205440	0
	Bolivia	2202996	0
	Brazil	158474256	0
	Canada	613322424	20
	Chile	21127380	0
	Colombia	21622020	0
	Dominican Republic	3383112	0
	Ecuador	3102984	0
	Egypt	1620000	0
	France	522978986	42
	Germany	103039776	36
	India	243642861	3
	Iran	28740393	0
	Iraq	12020151	0
	Israel	20596983	0
	Italy	78096843	4
	Kazakhstan	23800843	0
	Kuwait	5744200	0
	Mexico	11626065	0
	Oman	3401332	0
	Pakistan	22647592	0
	Panama	2650945	0
	Peru	27427719	0
	Philippines	18369956	3



Impacto de Casos Graves e Críticos nos Recursos Hospitalares

Avaliar como o número de casos graves e críticos afeta os recursos hospitalares pode informar a necessidade de expansão de capacidade ou reforço em determinadas áreas.

```
SELECT d.Data, SUM(f.Serios_Criticos) AS Casos_Serios_Criticos
```

```
FROM Fato_Covid19 f
```

```
JOIN dim_data d ON f.Id_data = d.Id_data
```

```
GROUP BY d.Data;
```

Result Grid		
	Data	Casos_Serios_Criticos
▶	2020-01-22 00:00:00	75066
	2020-01-23 00:00:00	75066
	2020-01-24 00:00:00	75066
	2020-01-25 00:00:00	75066
	2020-01-26 00:00:00	75066
	2020-01-27 00:00:00	75066
	2020-01-28 00:00:00	75066
	2020-01-29 00:00:00	75066
	2020-01-30 00:00:00	75066
	2020-01-31 00:00:00	75066
	2020-02-01 00:00:00	75066
	2020-02-02 00:00:00	41553