

Engenharia de dados aplicado à COVID 19

01.

DESENVOLVIMENTOS

Explicação e Funcionamento do back-end, Databricks e Front-end.

02.

EVOLUÇÃO DA SOLUÇÃO

Melhorias possíveis e lições aprendidas.

DESENVOLVIMENTOS



Back-End

Requisição de informações via API, verificação da informação e inserção no banco de dados



Databricks

Extração da informação do banco de dados, transformação, testes de hipóteses e insights

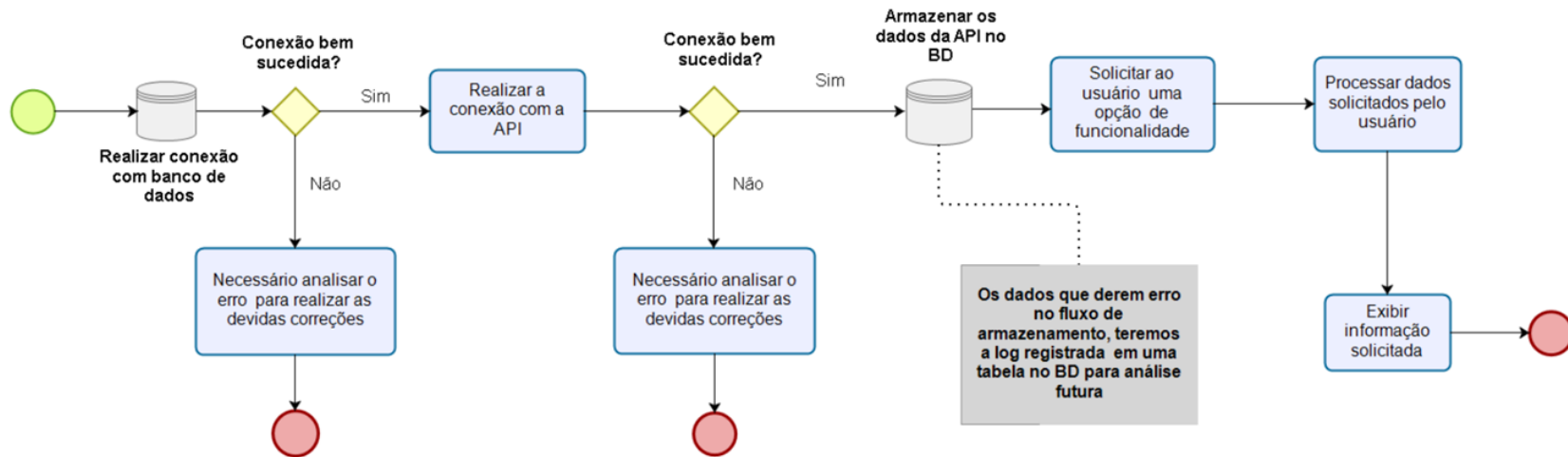


Power BI

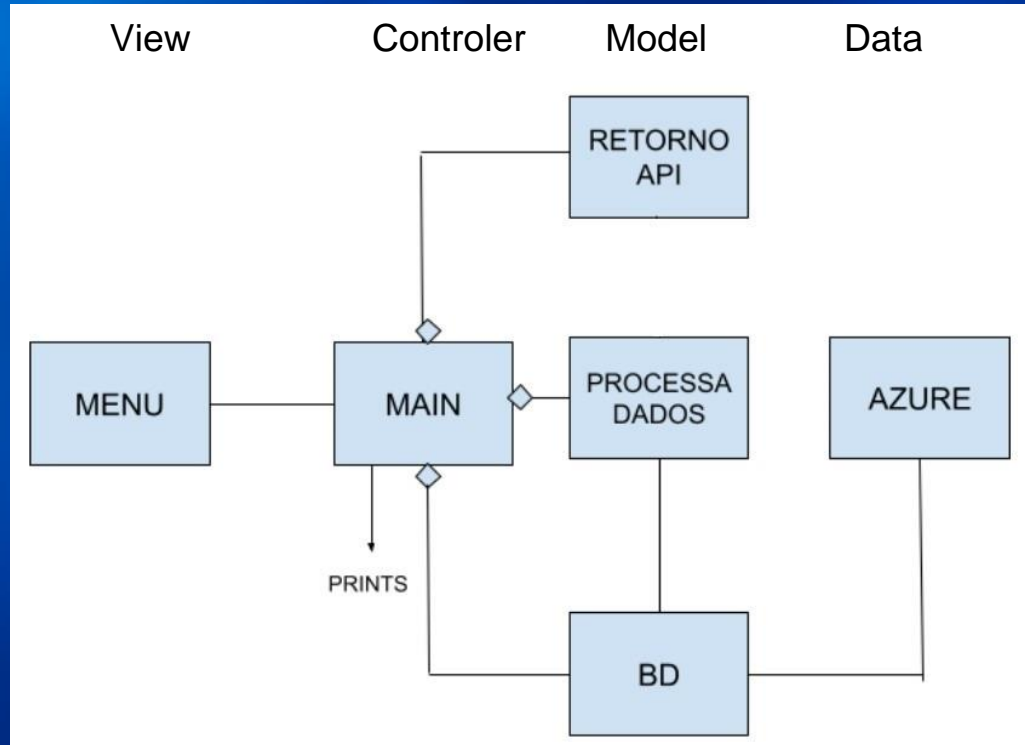
Dashboards com informações e indicadores dos casos de Covid no mundo/Brasil.

ARQUITETURA BACK-END

Funcionalidade software

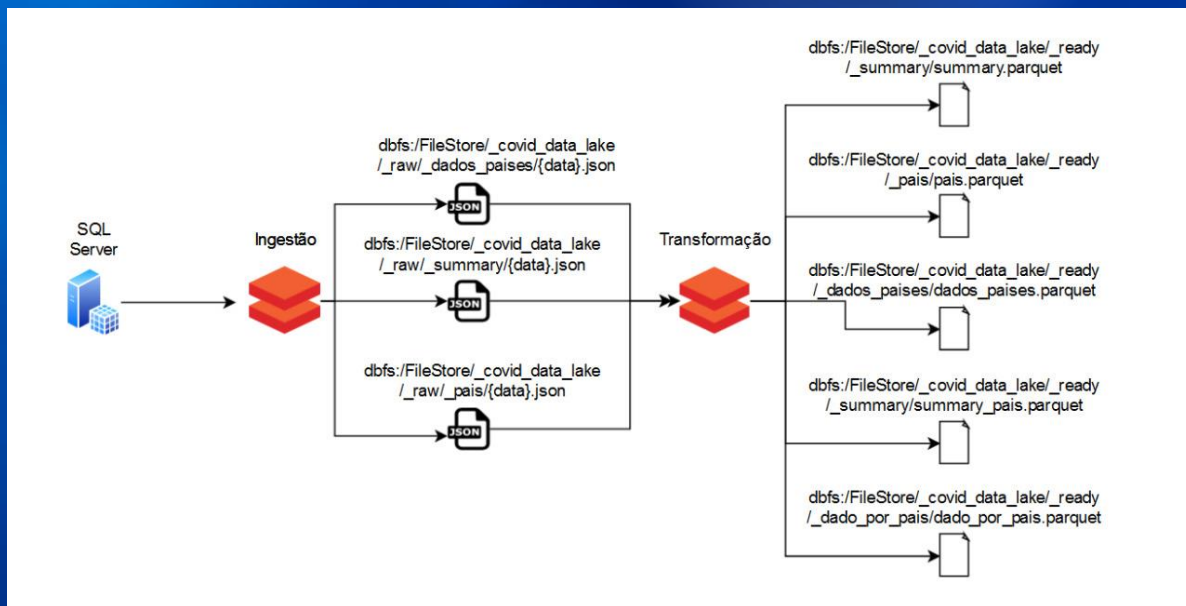


ESTRUTURA DO SOFTWARE - BACK-END



FUNCIONALIDADE FRONT-END

Divisão dos Notebooks



Resultado das consultas

22:19:14 Panorama diário de quantidade de casos confirmados de COVID-19 dos 10 países do mundo com maiores números.

	NOME	NEWCONFIRMED
0	Brazil	79876
1	United States of America	58915
2	France	29674
3	India	22854
4	Italy	22275
5	Germany	21163
6	Poland	17277
7	Czech Republic	15380
8	Turkey	14556
9	Russian Federation	8976

22:20:45 Total de casos confirmados por COVID-19 dos 10 países do mundo com maiores números.

	NOME	TOTALCONFIRMED
0	United States of America	29236735
1	India	11285561
2	Brazil	11202305
3	Russian Federation	4302726
4	United Kingdom	4247879
5	France	4022429
6	Spain	3172101
7	Italy	3123368
8	Turkey	2821943
9	Germany	2541781

22:19:31 Panorama diário de quantidade de mortes de COVID-19 dos 10 países do mundo com números.

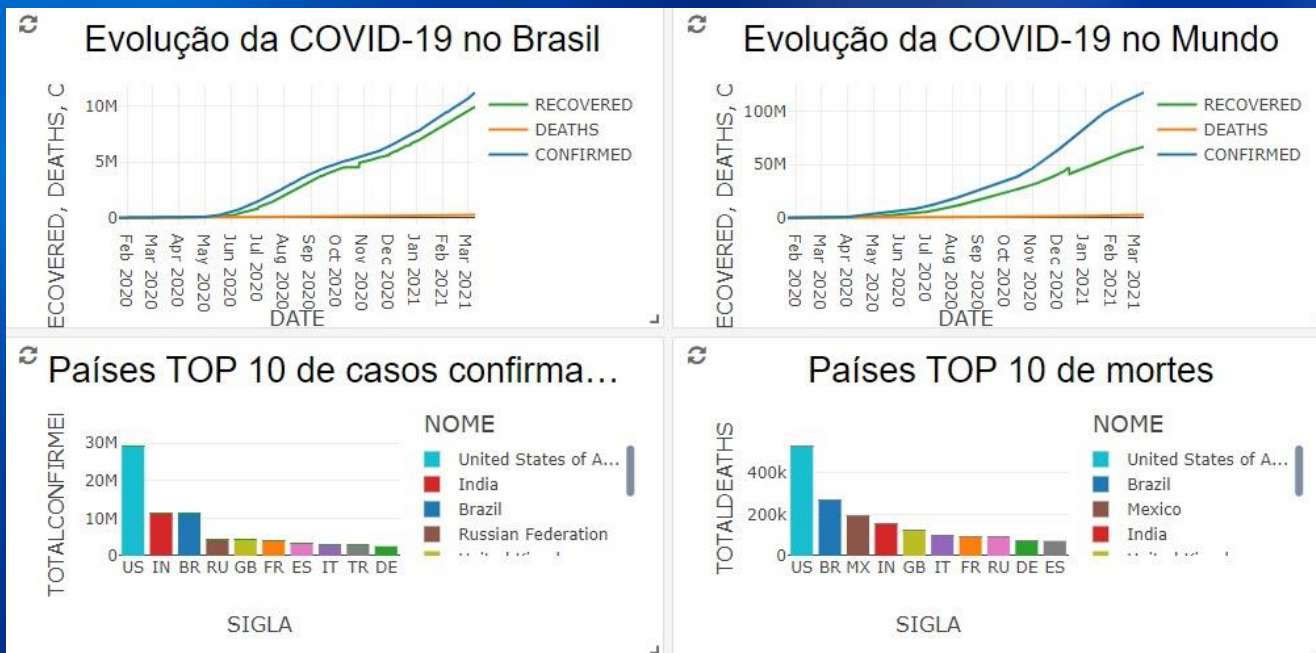
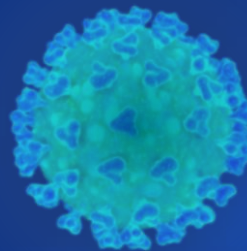
	NOME	NEWDEATHS
0	Brazil	2286
1	United States of America	1564
2	Mexico	702
3	Russian Federation	458
4	Poland	398
5	Italy	332
6	Germany	324
7	France	252
8	Czech Republic	238
9	Spain	234

22:20:26 Total de mortes por COVID-19 dos 10 países do mundo com maiores números.

	NOME	TOTALDEATHS
0	United States of America	529263
1	Brazil	270656
2	Mexico	192491
3	India	158189
4	United Kingdom	125222
5	Italy	100811
6	France	89707
7	Russian Federation	88773
8	Germany	72858
9	Spain	71961

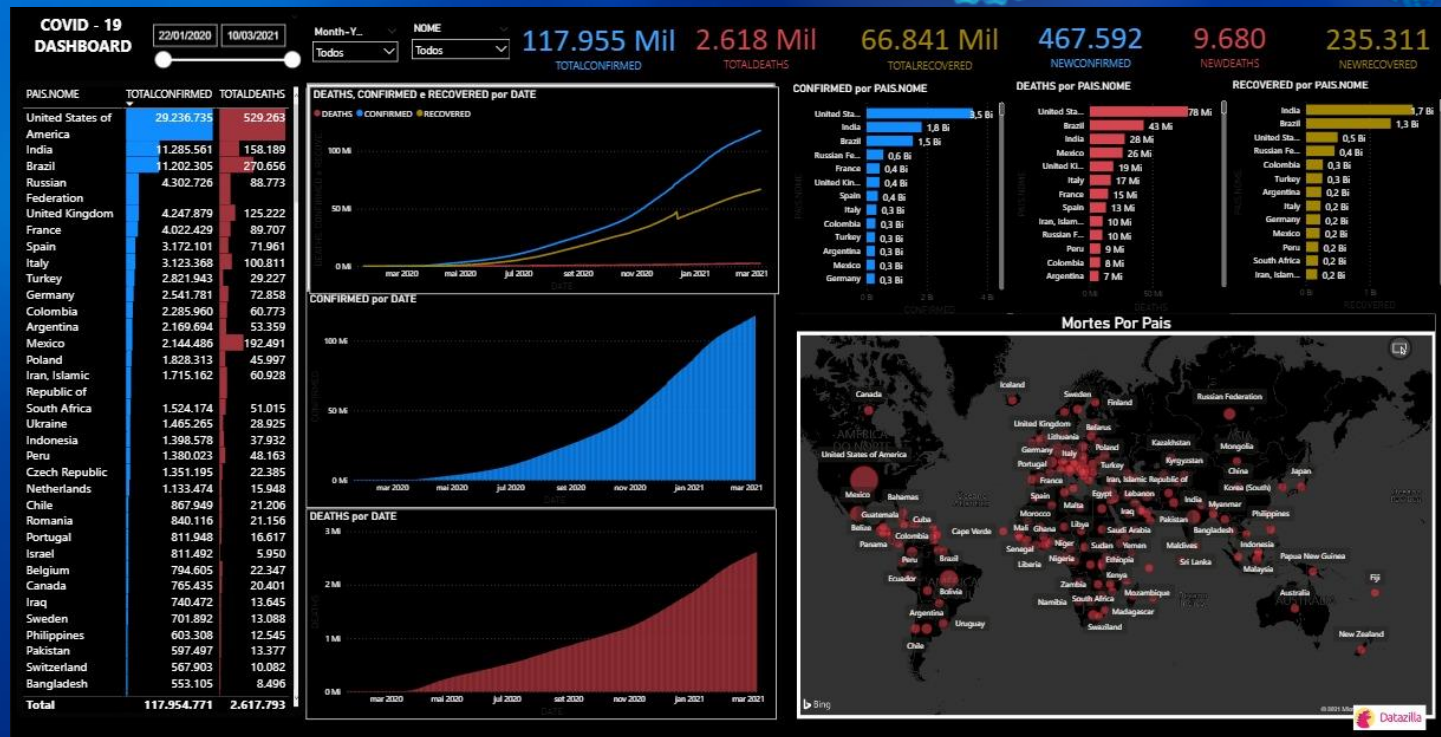
FUNCIONALIDADE FRONT-END

View DataBricks



FUNCIONALIDADE FRONT-END

Conexão com PowerBI



FUNCIONALIDADE FRONT-END

Conexão com PowerBI

ID	PAIS.NOME	TOTALCONFIRMED
562	United States of America	29.236.735
457	India	11.285.561
404	Brazil	11.202.305
521	Russian Federation	4.302.726
561	United Kingdom	4.247.879
440	France	4.022.429
541	Spain	3.172.101
463	Italy	3.123.368
557	Turkey	2.821.943
444	Germany	2.541.781

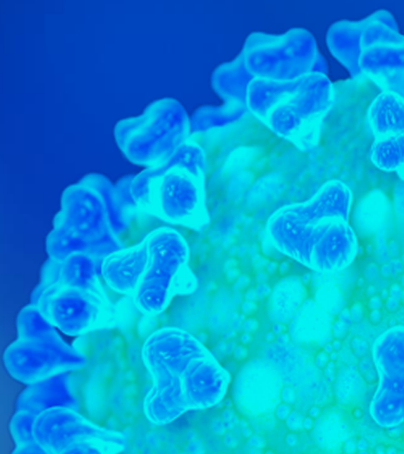
ID	PAIS.NOME	NEWCONFIRMED
404	Brazil	79.876
562	United States of America	58.915
440	France	29.674
457	India	22.854
463	Italy	22.275
444	Germany	21.163
516	Poland	17.277
425	Czech Republic	15.380
557	Turkey	14.556
521	Russian Federation	8.976

ID	PAIS.NOME	TOTALDEATHS
562	United States of America	529.263
404	Brazil	270.656
491	Mexico	192.491
457	India	158.189
561	United Kingdom	125.222
463	Italy	100.811
440	France	89.707
521	Russian Federation	88.773
444	Germany	72.858

ID	PAIS.NOME	NEWDEATHS
404	Brazil	2.286
562	United States of America	1.564
491	Mexico	702
521	Russian Federation	458
516	Poland	398
463	Italy	332
444	Germany	324
440	France	252
425	Czech Republic	238
541	Spain	234

DIFICULDADES E LIÇÕES APRENDIDAS



- API Instável – Foi necessário criar caches em disco e trabalhar extensivamente com exceções
 - Validar os tipos criando Schema de dados no DataFrame
 - Processar os dados antes de enviar verificando duplicidade através de Timestamp e Nomes dos Países, criando o relacionamento de chaves
 - Conexão de internet Instável forçou inserir linha a linha, fazendo commits depois de um número de inserts (ao invés de se utilizar funções otimizadas para muitos inserts)
 - Maior controle para o usuário, fragmentando as tarefas e possibilitando ao usuário fazê-las de acordo com a necessidade
 - Foco na Qualidade, Desenvolvimento para “outra” equipe, Boas Práticas de Design
- 

QA

#Case	Descrição	Paso	Resultado esperado	Resultado obtido
MENU-01	Verificar as opções no menu	Rodar o script python	<p>O seguinte menu e printado na tela:</p> <pre> "Digite a opção que deseja executar:\n" "===== Consulta =====\n" "1 - Exibi TODAS as consultas na Tela\n" "2 - Exibi Panorama diário de quantidade de casos confirmados de COVID-19 dos 10 países do mundo com maiores números\n" "3 - Panorama diário de quantidade de mortes de COVID-19 dos 10 países do mundo com números\n" "4 - Total de mortes por COVID-19 dos 10 países do mundo com maiores números\n" "5 - Total de casos confirmados por COVID-19 dos 10 países do mundo com maiores números\n" "===== Carga de Dados =====\n" "6 - Carregando Dados da API no buffer CSV\n" "7 - Carregar Todos os Dados da API nas Tabelas\n" "8 - Carregar Todos os Dados do BUFFER nas Tabelas\n" "9 - Carregar PAIS na Tabela\n" "10 - Carregar Dados de PAISES na Tabela\n" "11 - Carregar Sumario de PAISES na Tabela\n" "===== Manutenção =====\n" "12 - Cria estrutura de tabelas no DB Selecionado\n" "13 - Limpar conteúdo das Tabelas\n" "14 - Limpar conteúdo DADOS_PAISES\n" "15 - Limpar conteúdo SUMMARY_PAISES\n" "16 - Limpar conteúdo LOG\n" "===== FIM =====\n" "17 - Sair do Programa\n" </pre>	OK
MENU-02	Inserir um numero fora do rango das opcoes	<p>Rodar o script python</p> <p>Inserir um numero fora do rango (por ex -10)</p>	Se mostra uma mensagem de erro indicando que a opcao no e valida e pedindo para ingresar novamente um numero	OK
MENU-03	Inserir um caracter nao numerico	<p>Rodar o script python</p> <p>Inserir uma letra como opcao</p>	Se mostra uma mensagem de erro indicando que a opcao no e valida e pedindo para ingresar novamente um numero	OK
MENU-04	Validar qualquer das opcoes sem conexao a internet ou o servidor	<p>Rodar o script python</p> <p>Selecionar uma opcao de 1 a 7</p>	Se mostra uma mensagem indicando que nao ha conexao com o servidor, sem quebrar o programa	OK
MENU-05	Validar opcao de baixar o arquivo de dados covid	<p>Rodar o script python</p> <p>Selecionar a opcao 1</p>	O arquivo csv é descarregado com sucesso	OK
MENU-06	Validar opcao de criar tabelas no servidor	<p>Rodar o script python</p> <p>Selecionar a opcao 2</p> <p>Verificar as tabelas no servidor SQL</p>	As tabelas de PAIS, DADOS_PAISES, SUMMARY_PAISES e LOG estao criadas corretamente	OK



OBRIGADO!

GRUPO 2 - Datazilla

Deive Audieres Leal

Gabriel Ballesteros

George Razera

Judson Santana

Leonardo Moreno Giantin Rodrigues

Tiago Fernandes DAgostino

Viviane Jordão Nyitray

CREDITS: This presentation template was created by Slidesgo,
including icons by Flaticon, and infographics & images by Freepik