



**BlueShift**

**Business Documentation**

**São Paulo, 17 de dezembro de 2020**

CONTROLE DE VERSÃO			
Autor	Versão	Data	Descrição
Gustavo Henrique G.R	1.0	17/12/2020	Criação do documento.
Isabele Ruffo Lopes Gustavo Henrique G.R	1.1	17/12/2020	Ajustes nas informações
Isabele Ruffo Lopes Gustavo Henrique G.R	1.2	13/01/2020	Finalização do documento.

## 1. Introdução

Este documento tem como objetivo apresentar, em detalhes, os requisitos do Projeto Integrado do ponto de vista técnico, exigido pela empresa Blueshift como forma de treinamento do programa BlueShift Academy.

## 2. Solicitação

O objetivo do projeto é apresentar informações relacionadas ao Covid-19 ( número de mortos e infectados) e uma comparação dos gastos do governo entre os anos de 2019 e 2020, ressaltando o peso da situação pandêmica atual dentro do governo.

A entrega do projeto foi dividida em duas partes: a primeira se dá no consumo dos dados e repositórios, disponíveis nos portais públicos: Portal da Transparência e Brasil.io, através de um script em Python para download dos dados, um script em R para transformação dos dados e um script final para criação das tabelas no Databricks.

A segunda parte foca-se no desenvolvimento de dashboards interativos utilizando o *PowerBI*.

### 3. Premissas da solução

#### 3.1. Origem e especificação dos dados

Os dados foram retirados de repositórios públicos, de forma automatizada, nos formatos ZIP, GZIP e CSV, contando com 25 links diferentes, conforme mostrado abaixo:

- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/201901>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/201902>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/201903>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/201904>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/201905>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/201906>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/201907>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/201908>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/201909>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/201910>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/201911>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/201912>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/202001>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/202002>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/202003>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/202004>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/202005>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/202006>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/202007>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/202008>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/202009>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/202010>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/202011>
- <http://www.portaltransparencia.gov.br/download-de-dados/despesas-favorecidos/202012>
- [https://data.brasil.io/dataset/covid19/caso\\_full.csv.gz](https://data.brasil.io/dataset/covid19/caso_full.csv.gz)

### 3.2. Ambiente de desenvolvimento

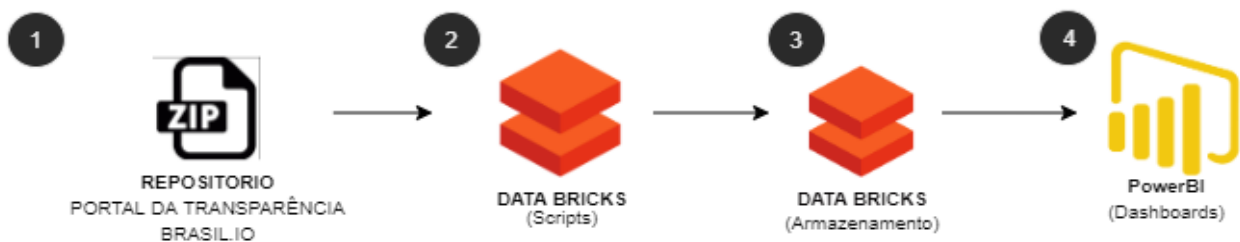
A empresa Blueshift disponibilizou todos os acessos e ferramentas necessárias para o desenvolvimento da arquitetura definida na apresentação do Blueprint.

Os ambientes de desenvolvimento utilizados para realização desse projeto foram:

- Azure Databricks Microsoft;
- Power BI;

### 4. Modelo da arquitetura sugerida

A figura abaixo apresenta a arquitetura da solução com base no levantamento de requisitos e entendimento do negócio.



1. Repositório público (Portal da transparência, Brasil.io);
2. Extração dos arquivos via **Databricks**, utilizando um script em Python;
3. Manipulação dos dados, via **Databricks**, utilizando scripts em Python, Linux e R;
4. Consumo das tabelas e montagem do Dashboards utilizando o **Power BI**.

### 5. Repositório

Abaixo, segue o link para o github do nosso projeto:

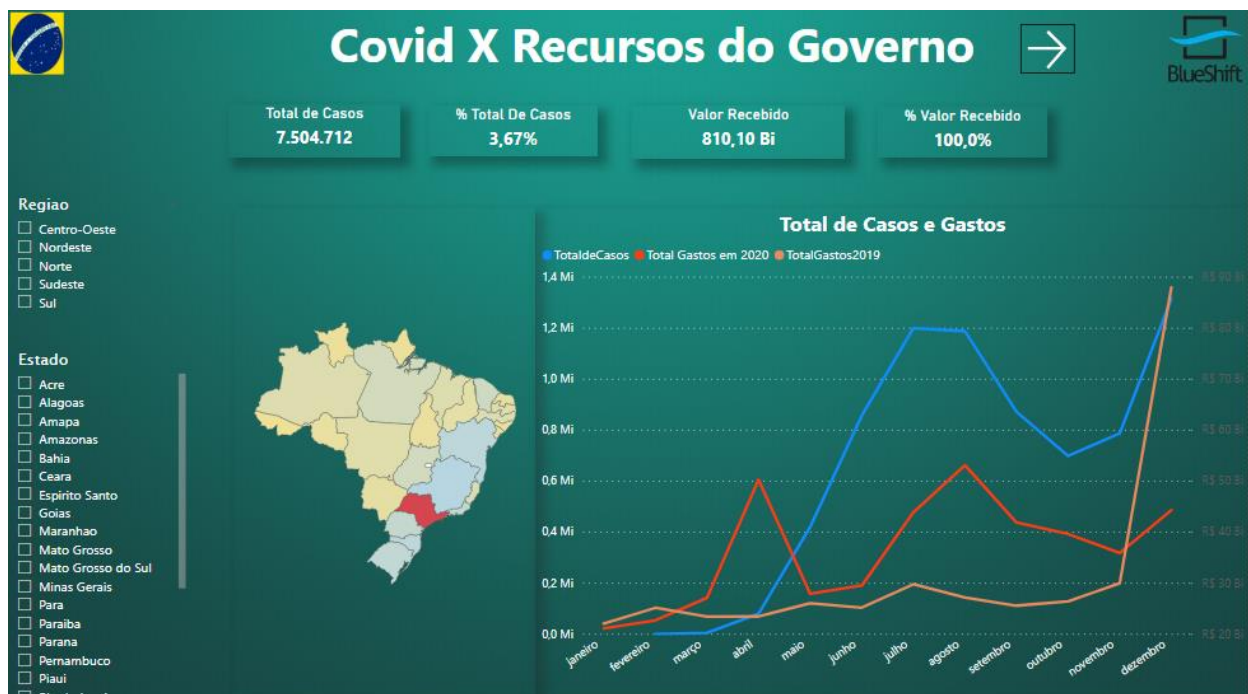
<https://github.com/GustavoGattas/ProjetoIntegrado>

## 6. Dashboards

### 6.1. Covid X Recursos do Governo

Baseado nas informações das bases de dados, obtivemos o seguinte quadro de resultados:

- O total de casos de Covid-19 em 2020, total de gastos do Governo ao longo do ano de 2019 e 2020, apresentado em meses.
- Cards destacando o total de casos e a sua porcentagem, o valor geral recebido e sua porcentagem.
- Um mapa do Brasil, com a opção de filtragem por Região e Estado.



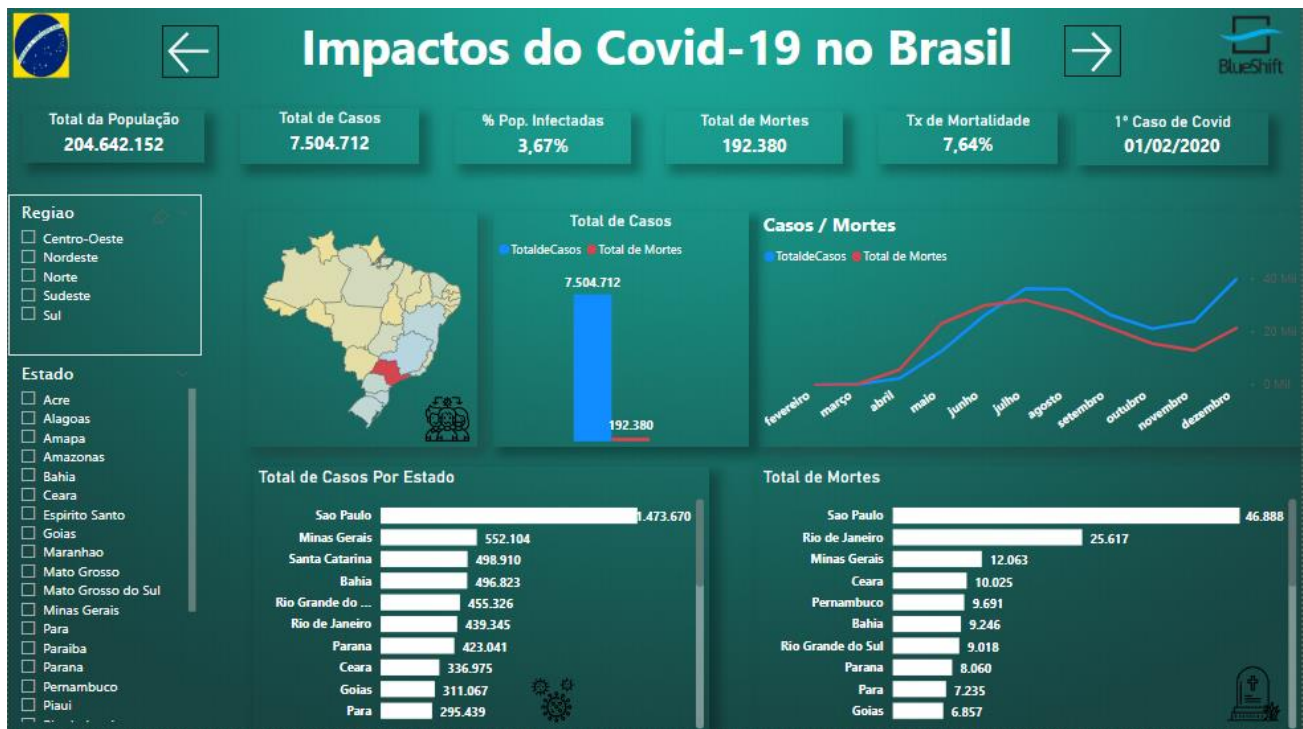
### 6.2. Impactos do Covid-19 no Brasil

Baseado nas informações das bases de dados, obtivemos o seguinte quadro de resultados:

- Um gráfico apresentando o total de casos por estado, destacando os 10 estados com mais casos.
- Um gráfico apresentando o total de mortes por estado, destacando os 10 estados com mais casos
- Comparação dos casos e mortes ao longo do ano de 2020, apresentado em meses.
- Cards destacando o total de casos e mortes, total da população, porcentagem

infectada, taxa de mortalidade e a data do primeiro caso registrado de Covid no Brasil.

- Comparação do total de casos com o total de mortes no Brasil.
- Um mapa do Brasil, com a opção de filtragem por Região e Estado.



### 6.3. Distribuição de recursos no Brasil

Baseado nas informações das bases de dados, obtivemos o seguinte quadro de resultados:

- Um comparativo dos recursos recebidos por estado no ano de 2019 e 2020.
- Um comparativo dos gastos do governo no ano de 2019 e 2020.
- Um gráfico de recursos recebidos por estado, destacando os 10 estados que mais receberam.
- Um gráfico de total de recursos por órgão governamental, realçando os 10 órgãos que mais gastaram ao longo de 2020.
- Cards destacando o total da população, total de gastos em 2019 e sua porcentagem, total de gastos em 2020 e sua porcentagem.
- Um mapa do Brasil, com a opção de filtragem por Região e Estado.



Os dashboards estão disponíveis no endereço abaixo:

<https://github.com/GustavoGattas/ProjetoIntegrado>