



Desafio Data Engineer

O objetivo deste desafio é analisar sua capacidade em pesquisar e desenvolver um pipeline de dados através da ferramenta [Apache Beam](#). Para isto, esperamos receber de você um script desenvolvido em **Python** e que seja capaz de executar todos os passos necessários para entregar os resultados solicitados .

Os arquivos necessários estão disponíveis através do arquivo compactado .zip [disponível para download neste link](#), no qual existem dois arquivos públicos que foram extraídos do site do IBGE e do Corona Vírus Brasil:

1- EstadosIBGE.csv – Informações gerais dos estados

2- HIST_PAINEL_COVIDBR_28set2020.csv – Dados históricos do Covid-19 no Brasil

Tarefa

Criar um pipeline de dados, utilizando do Apache Beam, que seja capaz de ler os dois arquivos em anexo e que o resultado desse pipeline sejam os dois seguintes arquivos:

1º arquivo

Agregado de informações por estado.

- Formato: CSV
- Informação: Regiao, Estado, UF, Governador, TotalCasos, TotalObitos

2º arquivo

Com base nos mesmos resultados gerados no 1º arquivo, gerar um arquivo .json (válido) onde cada coluna do arquivo anterior, seja uma key dentro desse json.

```
[
  {
    "Regiao": "Sudeste",
    "Estado": "Rio de Janeiro",
    "UF": "RJ",
    "Governador": "WILSON JOSÉ WITZEL",
    "TotalCasos": 99999,
    "TotalObitos": 99999,
  }
]
```

Entrega esperada

Código em um repositório público do GitHub, com o script desenvolvido junto aos dois arquivos gerados e um README.md explicando como foi desenvolvido o script e como o executar.

Serão considerados diferenciais:

- Utilização de boas práticas de desenvolvimento
- Versionamento contínuo do código no GitHub
- Documentação do código e do projeto
- Explicações sobre o porquê da utilização do Apache Beam, demonstrando domínio sobre o framework

Período de desenvolvimento

Você tem 7 dias corridos para elaborar esse desafio, a partir da data recebimento do mesmo.

Notificar a finalização por e-mail para paloma.pereira@hurb.com e lucas.rolim@hurb.com

Boa sorte! ;)