



Desafio Luizalabs

Esse é o desafio para **Engenharia de Dados** do Luizalabs. Esse desafio contém 2 partes onde você precisará demonstrar seus conhecimentos em:

- Apache Spark (Scala ou Python)
- Estruturas de Dados
- Programação

Ao final de cada task você deve exportar um CSV para que possamos conferir o seu resultado. Você deve também versionar o seu código no github/bitbucket/gitlab de modo que possamos avaliar o seu desenvolvimento. Inclua no repositório os CSVs gerados. Devemos conseguir rodar seu código e obter o mesmo resultado dos seus CSVs.

Task 1

Leia o arquivo de texto *wordcount.txt*, e conte as palavras que contém até 10 letras. Conte também quantas palavras com mais de 10 letras existem no texto.

Dataset: <https://storage.googleapis.com/luizalabs-hiring-test/wordcount.txt>

Exemplo do dataset final:

```
[('two', 2),  
 ('behold', 1),  
 ('itself', 3),  
 ('MAIORES QUE 10', 3)]
```

Task 2

Leia o arquivo *pedidos.csv* e agrupe todos os cliente que fizeram mais de 2 compras nos dias de black friday dos últimos três anos. Filtre todos os clientes que são menores de 30 anos e coloque numa lista TODOS os códigos de pedido e a data em que foram efetuados. Adicione também a idade do cliente.

Dataset: https://storage.googleapis.com/luizalabs-hiring-test/clientes_pedidos.csv

Exemplo do dataset final:

codigo_cliente	numero_pedidos	lista_pedidos	idade
3f94dfd5ba70668e52a6e6a9bf4ab575	3	[[d14c67dcf2264bf59efbdbbe0c8ace53, 2018-11-13], [3bbc5ac27ff22d9b900252ec0b6dab87, 2018-11-12]]	23
c4b7b6ae67331fb8db19bc93b9ecef7b	4	[[5c8b22171871e274252fb44d2f95b509, 2018-11-17], [4f62a0dba70ce6d2542aa5f0e893c835, 2018-11-21]]	22
e2da55d2313cc12628069addb252bd12	3	[[737f65d6712e177bde291926d0b840f0, 2018-11-10], [7d862341a7a973971ebd28a242397543, 2018-11-16]]	18

Boa sorte!