



# Desafío Técnico

## Parte 1 – SQL

Recomendamos utilizar o [SQLFiddle](#), PostgreSQL como engine de SQL. Pense em usar CTEs e Window Functions para lhe auxiliar

Assuma um banco PostgreSQL, timezone do server é UTC.

1. Dada a seguinte tabela com os campos:

Updated\_at: timestamp

Id: int64

Name: string

Age: int64

updated_at	id	name	age
2021-03-25 12:59:30 UTC	4	Joao Silva	28
2021-03-25 12:59:15 UTC	4	João Silva	28
2021-03-24 19:50:30 UTC	4	João Sil	28
2021-03-24 19:46:30 UTC	4	João Silva	28
2021-03-24 19:45:30 UTC	1	Marcos Santana	25
2021-03-25 19:44:30 UTC	1	Marcos Santana Santos	25
2021-03-24 19:44:30 UTC	3	Airton	60
2021-03-24 19:44:30 UTC	5	Aurélio	35
2021-03-24 19:44:30 UTC	6	Carushow	27
2021-03-24 19:44:30 UTC	7	Perlita	29
2021-03-28 00:00:30 UTC	7	Perlita	30
2021-03-24 19:44:30 UTC	2	Leonardo	26
2021-03-24 15:44:20 UTC	4	João Silva	28
2021-03-24 14:20:30 UTC	4	João Silva	28
2021-03-24 13:44:30 UTC	4	João Silva	28
2021-03-24 12:25:30 UTC	4	João Silva	28
2021-03-24 10:44:30 UTC	4	João Silva	22

Retorne somente as linhas mais atualizadas para cada ID, ignorando qualquer eventual mudança no caminho. (Entregue o código SQL e um print screen do resultado)

2. Dada as duas tabelas abaixo (vendas\_cardapio e itens\_cardapio), responda às seguintes perguntas (resposta e query utilizada):

**Tabela: vendas\_cardapio**

data	id_item	preco	qt_vendidas
09/06/2020	111	6.5	32
10/06/2020	200	7	22
10/06/2020	340	5.25	15
11/06/2020	111	8	50
11/06/2020	340	5.5	9
10/06/2020	111	6.5	20
11/06/2020	200	7	10

**Tabela: itens\_cardapio**

id_item	nome	gluten	media_preco
111	Café expresso	nao	6.5
112	Cookie de Laranja	sim	8.49
200	Suco de Abacaxi	nao	7
340	Pão de queijo	sim	5

- Qual foi o nome do produto que mais teve quantidades vendidas durante o período todo (dia 09 ao 11)?
- Qual foi o nome do produto que não teve alteração de preço durante os reajustes? Esse produto tem glúten?
- Houve algum produto que não vendeu?
- Qual foi o valor total de vendas de cada dia?

## **Parte 2 – Programação**

1. Inverter os valores após os valores inteiros (casas decimais)

### **Exemplo:**

Input 234.567

Output esperado 432.765

Imagine dois arquivos. O primeiro contém 5 milhões de linhas de ids, o segundo segundo arquivo contém 4 milhões de linhas de ids. Gere um arquivo resultante (da forma mais performática que você conhecer) da diferença entre o arquivo A e o arquivo B.

**Nota: a leitura do arquivo não é o mais importante. A lógica, a estrutura de dados e explicações do porque terão pesos maiores.**

### **Exemplo:**

Arquivo a:

KKKKKKKKKKKKKKKKKK

MM45ds4dssd5sd65dcrewd

BBBBBBBBBBBBBBBBBB

Hk45ds4dssZZsd65dcrewd

Arquivo b:

BBBBBBBBBBBBBBBBBB

Hk45ds4dsXXXsd65dcrewd

KKKKKKKKKKKKKKKKKK

Hk45ds4dssd5sdTT8crew

### **Output esperado do seu programa:**

MM45ds4dssd5sd65dcrewd

Hk45ds4dssZZsd65dcrewd

OBS: Lembre de nos enviar o seu código

### Parte 3 – Arquitetura (não necessário, é um diferencial)

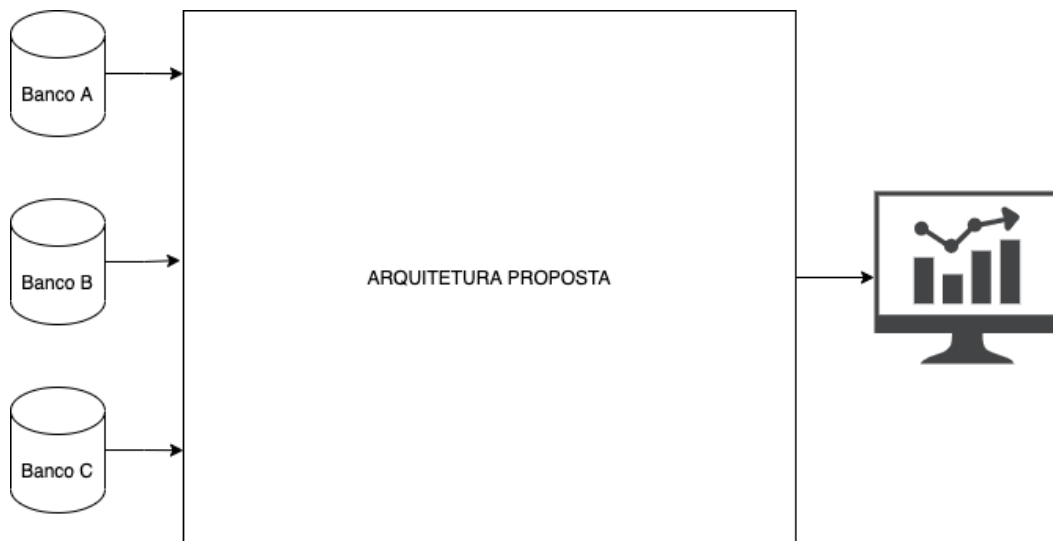
Queremos que você fale um pouco sobre a organização da arquitetura necessária para atender uma necessidade de negócio. Vamos falar de um caso de uso e gostaríamos que propusesse uma solução envolvendo as ferramentas com que tem familiaridade. Sinta-se à vontade para explicar de maneira descritiva, com desenhos ou qualquer forma que preferir, o importante mesmo é explicar quais são suas ideias.

Suponha que você tem três bancos de dados transacionais e tradicionais que serão sua fonte de dados. Eles possuem o mesmo schema mas são de organizações diferentes.

Seu trabalho é desenhar uma arquitetura que permita a união contínua e atualizada dessas bases em uma nova fonte de dados. Essa fonte de dados única será a entrada para dashboards da área de negócio.

Sinta-se à vontade para usar qualquer ferramenta, cloud provider, ou qualquer coisa que você julgar necessária.

O desenho abaixo ilustra o problema:



A entrega dessa etapa é o desenho da arquitetura com a explicação dos itens utilizados no pipeline.

**OBS: qualquer solução é bem vinda se estiver corretamente embasada e explicada.**