

Documentação do Projeto

Projeto Banco PAN

Sumário

0.0.	Atores.....	2
1.0.	Introdução.....	2
2.0.	Descrição do Projeto	2
2.1.	Nova Database em Nuvem.....	2
2.2.	Por que SQL Server?	3
2.3.	Nuvem.....	3
2.4.	Por que Cloud SQL?	3
2.5.	Despadronização e Validação dos Dados Inseridos	5
2.6.	ETL.....	7
3.0.	Conclusão.....	7
4.0.	The Interns	7

0.0. Atores

Atores	Funções
The Interns	<ul style="list-style-type: none">- Remodelar e construir a Database na nuvem- Construir APIs e REST APIs- Construir pipelines de ETL- Criar dashboards de análise de dados
CRM	<ul style="list-style-type: none">- Criar formulário de confirmação de dados- Tomar decisões com base em dados
Cliente	<ul style="list-style-type: none">- Inserir Dados- Contratar Produtos
Banco de Dados	<ul style="list-style-type: none">- Receber dados- Disponibilizar dados
APIs e REST APIs	<ul style="list-style-type: none">- Padronizar dados no Front-End- Validar dados no Back-end- Direcionar os dados para a Database- Direcionar Dados para ferramenta de ETL

1.0. Introdução

O projeto tem como finalidade reformular algumas instancias do Banco PAN visando uma nova base de dados confiável, permitindo assim a aplicação do **Business Intelligence** no negócio da instituição em questão. Isso permitirá a mesma tomar decisões mais assertivas, aumentando assim a satisfação de seus clientes paralelamente ao lucro do PAN.

2.0. Descrição do Projeto

2.1. Nova Database em Nuvem

Criação de uma nova base de dados **SQL Server** na nuvem, modelada de uma forma totalmente normalizada (Evitando assim redundâncias, inconsistências e possibilitando escalabilidades).

Modelo Lógico



Modelo Relacional



Para possibilitar a consulta de informações específicas dos produtos contratados pelos clientes, a chave de identificação do contrato e do produto específico será uma chave GUID (Um ID único global de 36 caracteres: “**ac9c1a9f-86cd-4ab0-a58e-5e40eac42587**”):

CONTRATO				SEGURO			
codigo_contrato	id_cliente	codigo_produto	data	codigo_seguro	codigo_tipo_seguro	validade	codigo_produto
ac9c1a9f-86cd-4a...	1	7	09/10/2022	ac9c1a9f-86cd-4a...	2	09/10/2022	7

2.2. Por que SQL Server?

Entre [MS SQL Server](#) e [Oracle SQL Developer](#), a escolha da primeira foi por conta do preço de sua licença [Enterprise Edition](#) (13.748 US\$), notoriamente menor que o da segunda (47.500 US\$). Além disso, o SQL Server apresenta as mesmas funções essenciais disponibilizadas pela Oracle:

- Proteção
- Suporte e Soluções de problemas
- Manutenção
- Backups

Observação: O Banco de Dados da Oracle possui algumas vantagens em relação ao SQL Server, porém ele é muito mais caro e não traz tantos benefícios que justifiquem seu preço.

2.3. Nuvem

A nuvem escolhida para a criação do Banco de Dados é a [Cloud SQL for SQL Server](#) (Google). Comparações entre as pioneiras da *Cloud Computing* foram feitas e essa, além de um excelente preço, atende a todas as necessidades do projeto para o Banco Pan, como:



- Disponibilidade
- Back-up e Disaster Recover
- Escalabilidade
- Segurança

Observação: Todas as outras atendem esses mesmos requisitos e até alguns a mais, porém seus preços são notoriamente maiores)

2.4. Por que Cloud SQL?

Comparações registradas que serviram de base para a escolha da mesma:




- Disaster Recovery:

			
Disponibilidade SLA	99.95%	99.99%	99.95%
Recuperação de Desastres	Automatico	Automatico	Automatico

• Disponibilidade:

			
Multi AZ	99.99%	99.99%	99.99%
Servidor Unico	99.5%	99.9% [Premium SSD] 99.5% [SSD]	90% [Hora em Hora]

• Segurança:

			
Firewall e VNET	✓	✓	✓
Criptografia em Transito	✓	✓	✓
Criptografia em Repouso	✓	✓	✓
Criptografia do lado do Cliente	✓	✓	✓
Autenticação CLOUD	RDS Proxy	Azure AD	Cloud SQL Proxy

• Escalabilidade:

			
CPU[vCore]	96	80	96
RAM[GB]	624	504	384
Armazenamento[TB]	30 [Auto Crescimento]	8 [Auto Crescimento]	16
Escalação Vertical tempo	5 Minutos	Segundos	Alguns Minutos

• Preço:

Amazon RDS for SQL Server 	
Instancias: 1	
Localização: São Paulo	
Total Horas: 730	
CPU: 8	Memória RAM: 32GB
Versão SQL Server: Enterprise Edition	
Armazenamento: 250 GB	
Backup: 250GB	
Preço Total Estimado: US\$ 5.608,00 por mês	

AZURE SQL Database 	
Instancias: 1	
Localização: Sul do Brasil	
Total Horas: 730	
CPU: 6	Memória RAM: Auto
Versão SQL Server: Enterprise Edition	
Armazenamento: 250 GB	
Backup: 250GB	
Preço Total Estimado: US\$ 4.425,39 por mês	

Cloud SQL for SQL Server 	
Instancias: 1	
Localização: São Paulo	
Total Horas: 730	
CPU: 8	Memória RAM: 52GB
Versão SQL Server: Enterprise Edition	
Armazenamento: 250 GB	
Backup: 250GB	
Preço Total Estimado: US\$ 3.184,57 por mês	

2.5. Despadronização e Validação dos Dados Inseridos

Para padronizar a forma como os dados são inseridos (evitando que usuários diferentes insiram a mesma categoria de informação de inúmeras maneiras), no *front-end* da aplicação do PAN, ocorrerá adição de máscaras para formatar os dados na hora de sua inserção. Para isso será utilizado **JavaScript** para manipular a página WEB e acessar cada campo do formulário.

Para a validação dos dados inseridos, no *back-end* da aplicação serão adicionados recursos para validar:

- Documentos:

→ **CPF**: A validade se dará pela verificação dos dígitos (**JavaScript**):

```
function TestaCPF(strCPF) {
    var Soma;
    var Resto;
    Soma = 0;
    if (strCPF == "00000000000") return false;

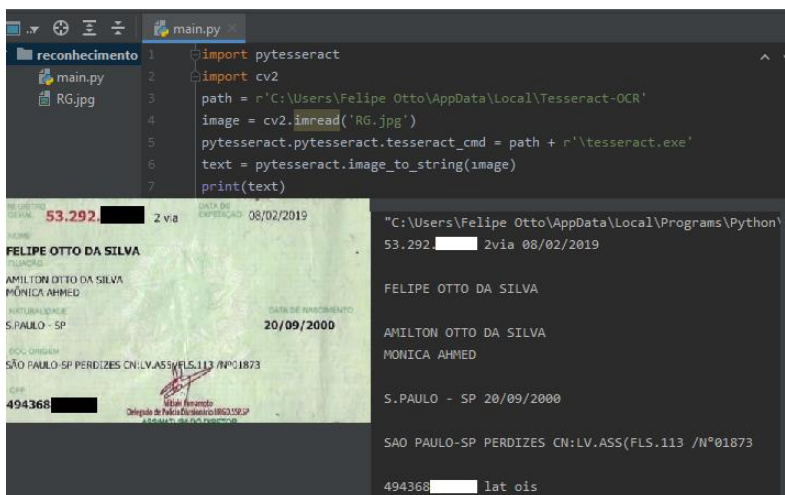
    for (i=1; i<=9; i++) Soma = Soma + parseInt(strCPF.substr(i-1, 1)) * (11 - i);
    Resto = (Soma * 10) % 11;

    if ((Resto == 10) || (Resto == 11)) Resto = 0;
    if (Resto != parseInt(strCPF.substr(9, 1)) ) return false;

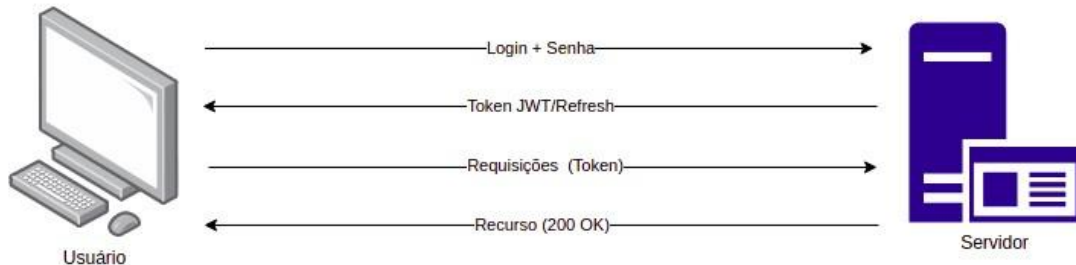
    Soma = 0;
    for (i = 1; i <= 10; i++) Soma = Soma + parseInt(strCPF.substr(i-1, 1)) * (12 - i);
    Resto = (Soma * 10) % 11;

    if ((Resto == 10) || (Resto == 11)) Resto = 0;
    if (Resto != parseInt(strCPF.substr(10, 1)) ) return false;
    return true;
}
```

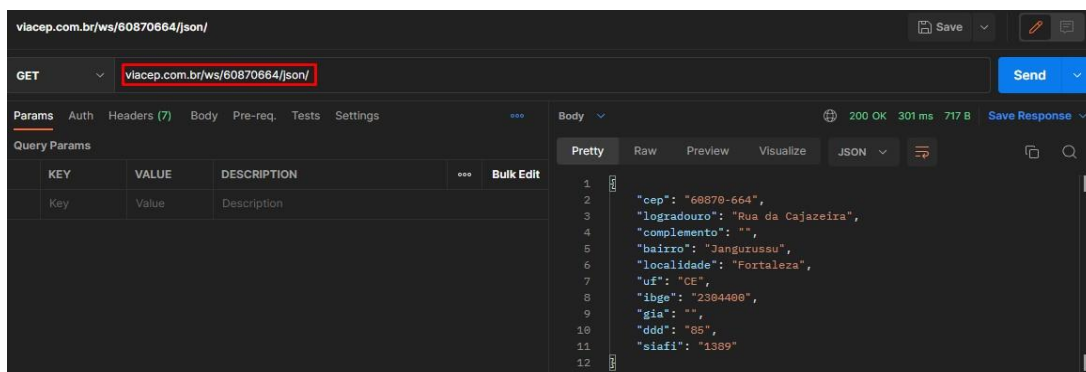
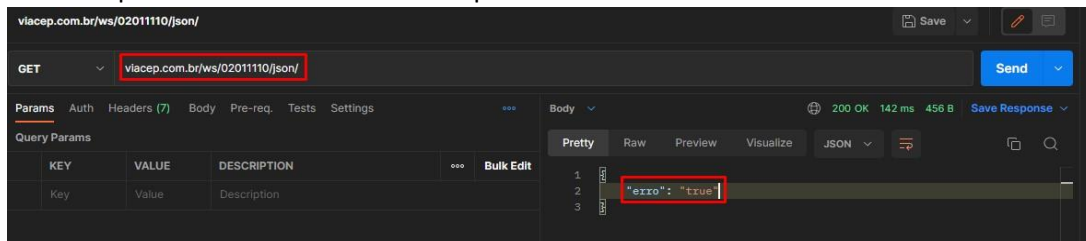
→ **RG e CNH**: A validade ocorrerá por reconhecimento de imagem (**Python**), que extrairá dados da imagem do documento e, antes de finalizar, os citará para o cliente confirmá-los:



- **Telefone e E-mail:** A validação ocorrerá via JWT (**JSON Web Token**). O servidor confirmará o login e senha e, a partir de então, será criado e enviado um token para o usuário que precisará digitá-lo para confirmar.
Observação: Essas informações são mantidas na máquina do cliente, e não no servidor (*Stateless*):



- **Endereço:** A consistência do endereço será garantida pela utilização da API **ViaCEP**, que valida um CEP inserido e retorna dados complementares sobre o mesmo, restando ao usuário apenas inserir o número e complemento:



Observação de Implementação: Órgãos federais e empresas privadas oferecem informações a respeito de pessoas (CPF, data de nascimento...). As duas melhores em questão de preço (pagamento ocorre conforme o consumo) e confiabilidade (pelos dados serem nativos da base da Receita Federal) são a **'SerPRO'** e **'SOA WebServices'**.

(<https://www.loja.serpro.gov.br/consultacpf>)

Por fim, os dados obtidos serão enviados para seus destinos (Banco de Dados e AWS Glue) via REST APIs, que são requisições que permite a transição de dados entre aplicações.

2.6. ETL

Finalizando o projeto, ocorrerá a integração de dados da base antiga para a nova através de um processo de ETL para que o Banco PAN não perca dados de seus clientes... Para isso, o CRM deverá criar um formulário de confirmação de dados que será enviado aos clientes cadastrados na base antiga via *push* (pela notificação do aplicativo). Assim, quando esses abrirem o aplicativo, deverão obrigatoriamente respondê-lo para continuar navegando.

Após respondido, os dados serão registrados na nova base via REST API e enviados para a AWS Glue que, através de *pipelines*, irá usá-los para consultar os produtos contratados pelo cliente e suas informações, extraíndo-os, transformando-os para que se adequem a nova regra de negócio e, enfim, os carregando para a nova *database*. Após esse processo, o registro manipulado será excluído da base antiga, permitindo aos poucos a amenização da infraestrutura do servidor *on-premises* do Banco PAN.

Esse processo ocorrerá via [Amazon Glue Data Brew](#), pois sua cobrança é feita por hora de execução e não por execução de pipeline como na [Azure Data Factory](#). Seu custo é de 69 centavos de dólar a hora (com 4 CPUs e 16 GB de memória).

Observação: Esse processo de ETL ocorrerá em batch, durante período noturno, horário de menor pico de uso de clientes (Das 22 as 9 em média).

3.0. Conclusão

Agora com todos esses processos feitos, o Banco PAN terá uma base totalmente confiável para a aplicação do Business Intelligence. Dashboards de dados serão criadas para que a instituição possa visualizar e, com base em dados, tomar a melhor decisão para eles.

4.0. The Interns

- Felipe Otto da Silva - (89108)
- Pedro Argentati - (88246)
- Pedro Henrique Castanheira - (89199)
- Rafael Tannous - (87486)
- Yasmin Verderame - (87023)