

## Documentação

A tomada de decisão baseada em dados é essencial para as empresas, mas decisões tomadas a partir de dados não confiáveis podem levar a perdas imensuráveis, criando desequilíbrios difíceis de reverter.

A qualidade dos dados depende de como esses elementos são preparados, inseridos, manipulados e armazenados. Na circunstância atual, infelizmente, o banco PAN sofre com esse ponto, isso por conta de três fatores:

1. O mal planejamento da infraestrutura de seu Banco de Dados.
2. A falta de máscaras para padronizar os dados inseridos.
3. A falta de validação dos dados inseridos.

## Nova Base de Dados

“A primeira etapa de nossa proposta será criar uma nova base de dados SQL Server na nuvem. Modelamos sua estrutura tornando-a totalmente normalizada, evitando assim redundâncias, inconsistências e possibilitando futuros crescimentos. Utilizamos chaves GUID (Um ID único global de 36 caracteres) que será utilizado para consultar informações específicas dos produtos contratados pelos clientes.

6E5F05EF-31C1-4936-8062-8DDB12B6DD92

Na escolha do SGBD ficamos entre o da Oracle e o MS SQL. O ponto crucial para nossa tomada de decisão foi o fato de a licença da Microsoft ser notoriamente mais acessível que a da Oracle.



vs



Além do preço acessível, a Enterprise Edition da MS apresenta as mesmas funções essenciais disponibilizadas pela Oracle:

Proteção: Logins autenticados no nível da instância e no nível do banco de dados.

Suporte e Soluções de problemas:

Manutenção: Tabelas geralmente armazenadas organizadas por índice

Backups:

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 assim que justifiquem seu preço.

## Cloud

Como citado, iremos criar uma nova database já em nuvem e o serviço que utilizaremos é o Cloud SQL for SQL Server da Google. Comparamos as pioneiras da Cloud Computing e, além de ter o preço mais acessível, atende a todas as necessidades que o banco Pan necessita, como:

Disponibilidade  
Back-up e Disaster Recover  
Escalabilidade  
Segurança

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

## Azure

▼ Banco de Dados SQL do Azure

Banco de dados único, vCore, Armazenamento de B...

Adiantado: R\$ 0,00

Mensal: R\$ 26.167,29

Suporte

SUPORTE:

Incluso

R\$ 0,00

PROGRAMA DE LICENCIAMENTO:

MCA (Contrato de Cliente da Microsoft)

MOSTRAR PREÇOS DE DESENVOLVIMENTO/TESTE

Custo adiantado estimado

R\$ 0,00

Custo mensal estimado

R\$ 26.167,29

## AWS

Utilização (somente sob demanda)  
Com a utilização, você ainda precisa interromper a instância para obter o custo-benefício. A utilização afeta apenas o preço do OnDemand para instâncias, não o armazenamento, backups etc.

100

%Utilized/Month

Modelo de definição de preço

OnDemand

Opção de implantação

Multi-AZ

Licença

License included

Database Edition

Standard

Configure Amazon RDS for SQL server

Informações

×

Armazenamento para cada instância do RDS

SSD de uso geral (gp2)

Quantidade de armazenamento

50

GB

► Mostrar cálculos

► Armazenamento de backup

Informações

Custo inicial total: 0.00 USD

Custo mensal total: 4.160,90 USD

Mostrar detalhes ▼

Salvar e visualizar resumo

Salvar e adicionar serviço

## GCP

### Cloud SQL for SQL Server

DB-HIGHMEM-8



Number of instances: 1

Location: Sao Paulo

Total hours: 730

Instance type: DB-HIGHMEM-8

USD 760.66

Database Version: ENTERPRISE

USD 2,330.16

Storage: 250 GiB

USD 63.75

Backup: 250 GiB

USD 30.00

**USD 3,184.57**

**Total Estimated Cost: USD 3,184.57 per 1 month**

Estimate Currency

USD - US Dollar



EMAIL

COPY SAVED URL

DOWNLOAD\*

16573,79

=

## Padronização dos dados

Agora, focando na despadronização dos dados (quando os usuários inserem o mesmo tipo de informação de maneiras distintas), iremos fazer uma validação no front end da aplicação, utilizando máscaras, para que essas cheguem no back-end já normalizados.

Esse processo será feito via JavaScript, manipulando o DOM (a página WEB) para acessar cada campo do formulário, com o objetivo de deixá-los padronizados.

Já no back-end da aplicação, verificaremos a veracidade dos dados inseridos, para combater fraudes e erros. Tá, beleza mas como faremos isso?

---

### Benefícios e Vantagens



Melhora a qualidade da avaliação de crédito



Minimiza o risco de fraudes



Segurança da informação



Informações sempre atualizadas



Acesso a informações diretamente do Governo, sem intermediários



Confiabilidade das informações



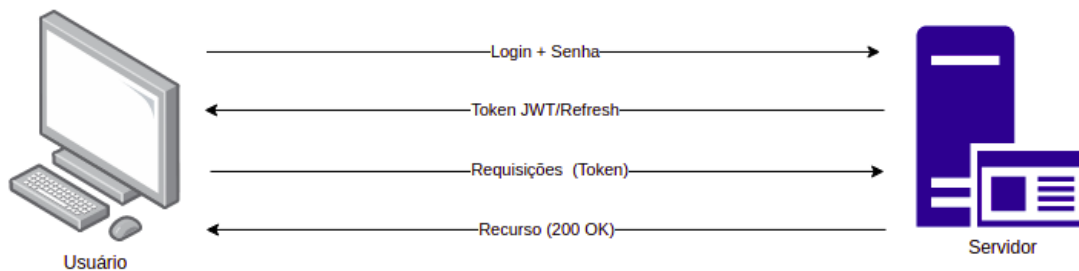
Possibilita checagem automática de informações

Tratando dos documentos, começando pelo CPF, verificaremos sua validade através de seus dígitos. Já outros documentos (RG e CNH) serão validados através de fotos tiradas pelo próprio cliente feitas com **reconhecimento de imagem**.

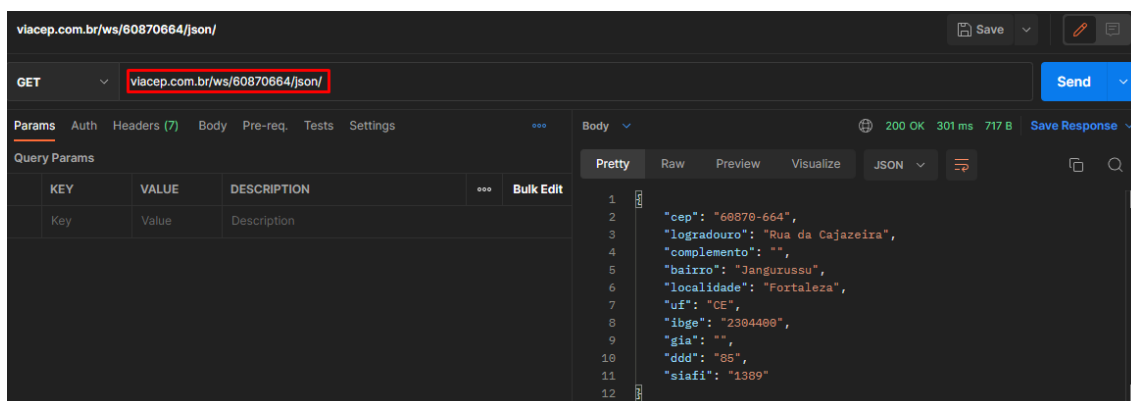
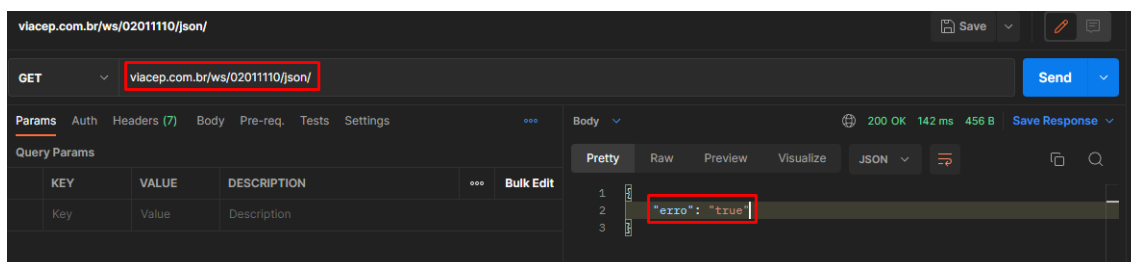
Isso será feito por meio de uma inteligência artificial criada com Python, que receberá o documento escolhido pelo cliente, extrairá os dados desses, e cadastrará no banco. Só que, antes disso, será exposto para o usuário um formulário com todos os dados inseridos para serem confirmados.

Em relação ao e-mail e telefone, utilizaremos o método de autenticação JWT (JSON Web Token). Após ter um login e senha validados pelo servidor, será criado um token que o usuário receberá em resposta e que permitirá o acesso a algum recurso. Essas informações são mantidas na máquina do cliente, e não no servidor (*Stateless*).

Esse token será enviado por email ou telefone da pessoa. Dependendo do caso, o usuário terá que inserir o código recebido na hora de finalizar o cadastro. Na parte das regras de negócio, será confirmado se o token inserido é igual ao gerado pela aplicação.



Por último, no endereço, utilizaremos a API 'ViaCEP' que valida o CEP e, caso exista, retorna dados complementares do mesmo, restando ao cliente apenas inserir o número e complemento.



Órgãos federais e empresas privadas oferecem informações a respeito das pessoas, como CPF, data de nascimento e afins. Através de pesquisas selecionamos as duas melhores em questão de preço (pagamento ocorre conforme o consumo) e confiabilidade (pelos dados serem nativos da base da Receita Federal): 'SerPRO' e 'SOA WebServices'.

Por fim, para enviar esses dados para o banco de dados, usaremos REST API's, que são requisições que permitem a comunicação de dados entre aplicações.

Aqui, podemos ter uma base dos valores por consulta do provedor SERPRO:

Link para mais informações -> <https://www.loja.serpro.gov.br/consultacpf>

## Processo de ETL

Por fim, realizaremos um ETL, para que o PAN não perca dados de seus clientes...

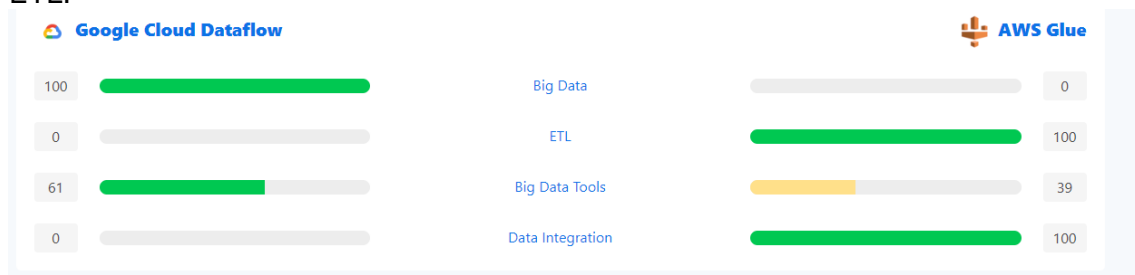
Para isso, primeiro criaremos um formulário de confirmação de dados e o enviaremos via push (pelo próprio app) para todos os clientes cadastrados na base antiga, assim, quando eles abrirem o aplicativo, deverão obrigatoriamente respondê-lo para continuar navegando.

Após respondido, os dados serão registrados na nova base via REST API e enviados para o **software de ETL**, que irá usá-los para consultar os produtos contratados pelo cliente e suas informações.

Após a consulta, caso exista, irá extraí-los e os transformará para que se adequem à nova regra de negócio, podendo assim finalmente carregá-los para a nova base. Por fim, todos os registros vinculados a aquele cliente serão excluídos da base antiga. Esse último processo irá permitir a amenização da infraestrutura do servidor físico do PAN.

Para realizar esse processo utilizaremos o 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).

Também procuramos a solução da google, o Dataflow, mas logo percebemos que o Glue é mais utilizado para o que estamos preocupados nesta etapa, o processo de ETL.



Esse processo de ETL ocorrerá em batch, durante o menor período pico de uso de clientes entre as 22 e as 9), totalizando cerca de 172R\$ mensais a cotação dolar.

Com nossa estrutura já pronta, pensamos em qual momento executar este processo para o carregamento da base nova, e o melhor horário para execução será em horários fora de pico e horários de pouco uso do aplicativo e site.

Com o decorrer do tempo, iremos aos poucos poder diminuir a infraestrutura do servidor físico, amenizando e muito o trabalho do banco PAN...

Agora com todos esse processos feitos, o banco PAN terá uma base totalmente confiável para a aplicação do Business Intelligence, aumentando assim o lucro e as tomadas de decisões da instituição junto ao contentamento de seus clientes.