

Process e Solution Design

Case Técnico RPA

Módulo C – Ingestão de E-mails & Anexos

Módulo D – Extração de Dados de PDF

1. PDD – Process Definition Document

1.1. Visão Geral do Processo

Este documento descreve a automação do processo de **ingestão de e-mails, tratamento de anexos PDF e extração estruturada de dados**, utilizada pela área de Operações e Compliance para consolidação diária de informações sensíveis recebidas por e-mail.

A automação substitui um fluxo manual altamente suscetível a erros por uma solução integrada, rastreável e escalável, desenvolvida em **Python**, onde os **Módulos C e D operam de forma encadeada** sob um único orquestrador.

1.2. Processo AS-IS

Atualmente, o processo é executado manualmente por analistas, conforme descrito abaixo:

1. Acesso manual à caixa de e-mail corporativa.
2. Identificação visual de e-mails com o assunto “*Relatório Diário*”.
3. Download manual dos anexos PDF.
4. Organização dos arquivos em pastas locais.
5. Abertura individual de cada PDF.
6. Localização manual de CPF e CEP.
7. Digitação das informações em planilhas Excel.
8. Validação manual (quando realizada).
9. Resposta manual ao remetente confirmando recebimento.

Tempo médio diário: 4 a 6 horas

Principais erros identificados:

- Erros de digitação e formatação;
- Falta de validação de CPF e CEP;
- Processamento duplicado de anexos;
- PDFs inválidos processados incorretamente;
- Ausência de rastreabilidade e histórico confiável

1.3. Processo TO-BE

Com a automação, o processo passa a ser executado de forma não assistida:

- Leitura automática da caixa de e-mail via API
- Identificação, validação e organização dos anexos
- Extração e validação automática de dados
- Consolidação em arquivo Excel estruturado
- Comunicação automática com os remetentes

Tempo estimado após automação: 5 a 10 minutos por execução.

1.4. Escopo da Automação

Incluído no escopo:

- Leitura automática da caixa de e-mail
- Tratamento e validação de anexos PDF
- Prevenção de duplicidade
- Extração de CPF e CEP
- Validação de formato e dígitos
- Consolidação em arquivo Excel
- Resposta automática ao remetente

Fora do escopo:

- Correção manual de PDFs inválidos
- Alteração do conteúdo dos documentos recebidos
- Integração direta com sistemas legados

1.5. Benefícios Esperados

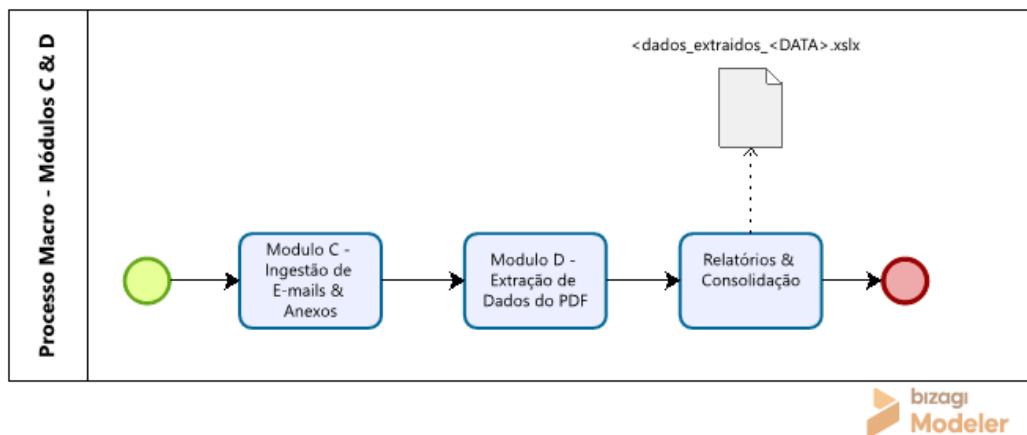
- Redução de tempo de até **80%**
- Eliminação de erros de digitação
- Validação automática de dados sensíveis
- Controle de duplicidade
- Organização padronizada de arquivos

- Dados estruturados para auditoria
- Processo escalável e resiliente

2. SDD – Solution Design Document

2.1. Arquitetura da Solução

A solução foi desenvolvida em **Python**, com arquitetura modular, onde os módulos C e D são executados de forma integrada sob um único fluxo orquestrado.



2.2. Ferramentas e Componentes

- **Linguagem:** Python
- **Integração de E-mail:** Gmail API
- **Manipulação de PDFs:** Bibliotecas Python (extração de texto)
- **Regex:** Identificação de CPF e CEP
- **Output:** Excel(.xlsx)

2.3. Módulo C – Ingestão de E-mails & Anexos

Responsabilidades:

- Conectar à caixa de e-mail via Gmail API
- Identificar mensagens com assunto “**Relatório Diário**”
- Validar existência e integridade de anexos PDF
- Salvar PDFs válidos em: *inbox/valid/YYYY-MM-DD*

- Evitar processamento duplicado de e-mails ou anexos
- Responder automaticamente ao remetente confirmando recebimento
- Tratar exceções:
 - E-mails sem anexo
 - PDFs corrompidos
- Mover e-mails inválidos para: *inbox/rejected*

2.4. Módulo D – Extração de Dados de PDF

Responsabilidades:

- Processar PDFs válidos recebidos do Módulo C
- Extrair automaticamente:
 - CPF
 - CEP
- Utilizar expressões regulares (regex) para identificação
- Validar:
 - Formato do CPF
 - Dígitos verificadores do CPF
 - Formato do CEP
- Consolidar resultados no arquivo: *dados_extraidos_<DATA>.xlsx*

Estrutura do arquivo:

arquivo	cpf	cpf_valido	cep	cep_valido	erro
---------	-----	------------	-----	------------	------

2.5. Estrutura do código

main.py	email_actions.py	pdf_actions.py	excel_actions.py
Orquestra todo o fluxo	Conexão com Gmail API	Leitura e validação de PDFs	Criação e atualização do arquivo Excel
Controla execução sequencial dos módulos	Leitura, filtro e resposta de e-mails	Extração de dados via Regex	Consolidação dos dados extraídos
Gerencia logs e exceções	Controle de duplicidade		

2.6. Tratamento de Exceções

- **Business Exceptions**
 - CPF ou CEP não identificado
 - PDF inválido ou incompleto
 - **System Exceptions**
 - Falha de conexão com API de e-mail
 - Erro na leitura de arquivos
- Todas as exceções são registradas em log para auditoria e suporte

3. Conclusão

A automação dos Módulos C e D elimina um processo manual crítico, reduzindo drasticamente o esforço operacional e os riscos associados à manipulação de dados sensíveis. A solução em Python, com arquitetura modular e integração via API, garante confiabilidade, escalabilidade e aderência às boas práticas corporativas.