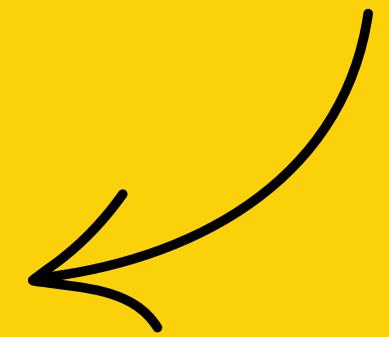


Pipeline de Dados: do zero ao estrelato

PROJETO SOULCODE & DM9





Álevy Bruno



Dionísio Ribeiro



Francisco Matheus



Joshua Bertocchi



Kleiton Santos



Lorena Gutierrez



Lucas Aguiar



Nicolly Rodrigues



Douglas Ribeiro



O problema identificado

O setor de Recursos Humanos (RH) enfrenta dificuldades na coleta, organização e análise de dados devido à **fragmentação dos dados**. Atualmente, os dados são **recebidos por e-mail** e consolidados em uma planilha, **sem um padrão definido**



Objetivos do projeto

01 Automação do processo

Eliminar a manipulação manual de arquivos e reduzir erros na coleta de informações

02 Estruturação e normalização

Transformar dados desorganizados e heterogêneos (diferentes fontes e colunas) em um modelo padronizado

03 Arquitetura escalável

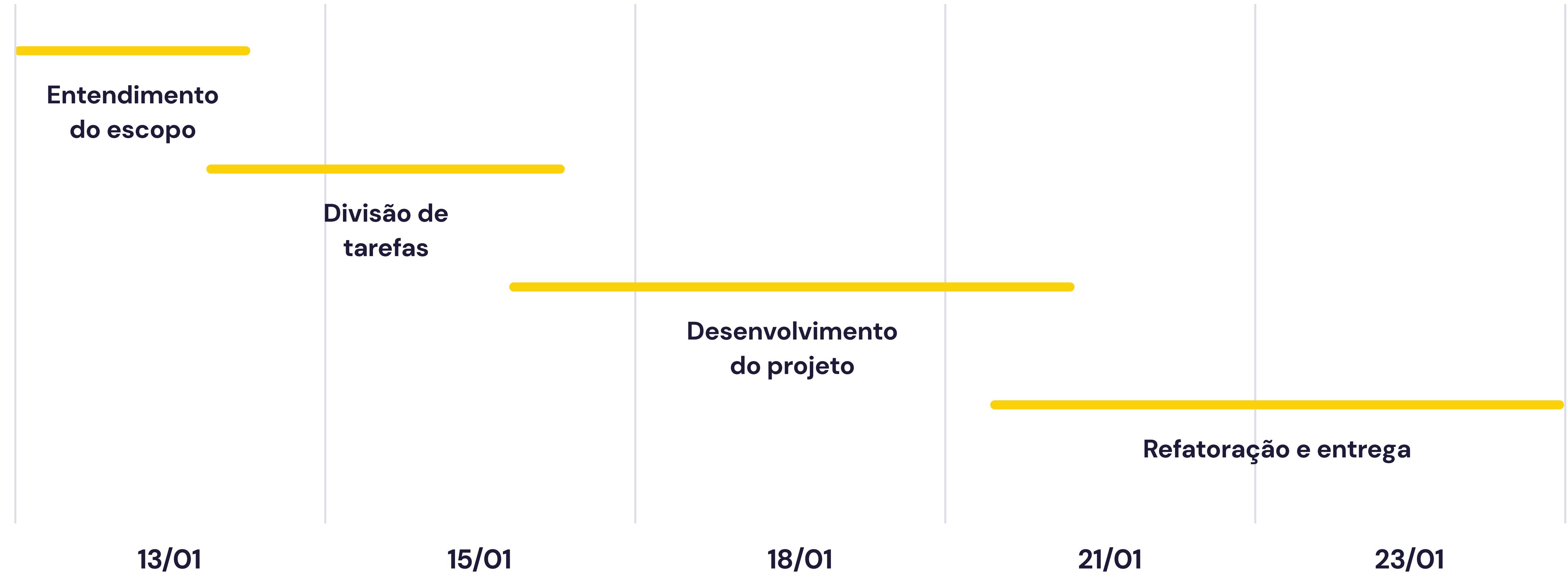
Reducir dependência de planilhas manuais, implementando uma pipeline persistente

04 Geração de insights confiáveis

Criar dashboards interativos para tomada de decisão depois de um processo de tratamento rigoroso



Roadmap



— O fluxo



Entendimento do Escopo

01

Alinhamentos e reuniões

Estes momentos foram importantes para alinhar com o professor e com o cliente sobre os desafios que enfrentaríamos

02

Diagrama de Fluxo de Dados

Nesta etapa, o DFD de nível Zero, ajudou a descrever o problema que a equipe iria resolver ao focar em como seria toda a trajetória dos dados

03

Prototipagem

O protótipo tinha como principal objetivo apresentar ao cliente a saída pretendida do sistema, sem ênfase no formato, mas sim no conteúdo



Etapa 1 

Extração

Etapa 3 

Visualização de Dados

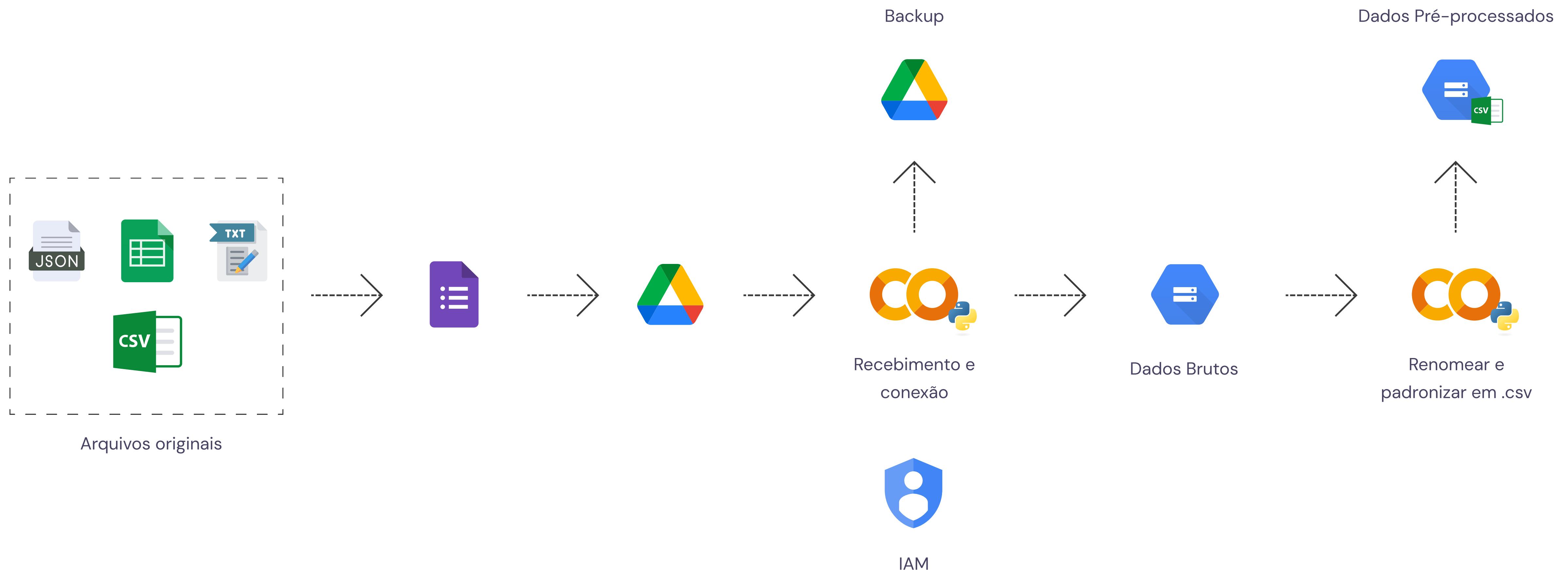
Etapa 2 

Tratamento e Transformação

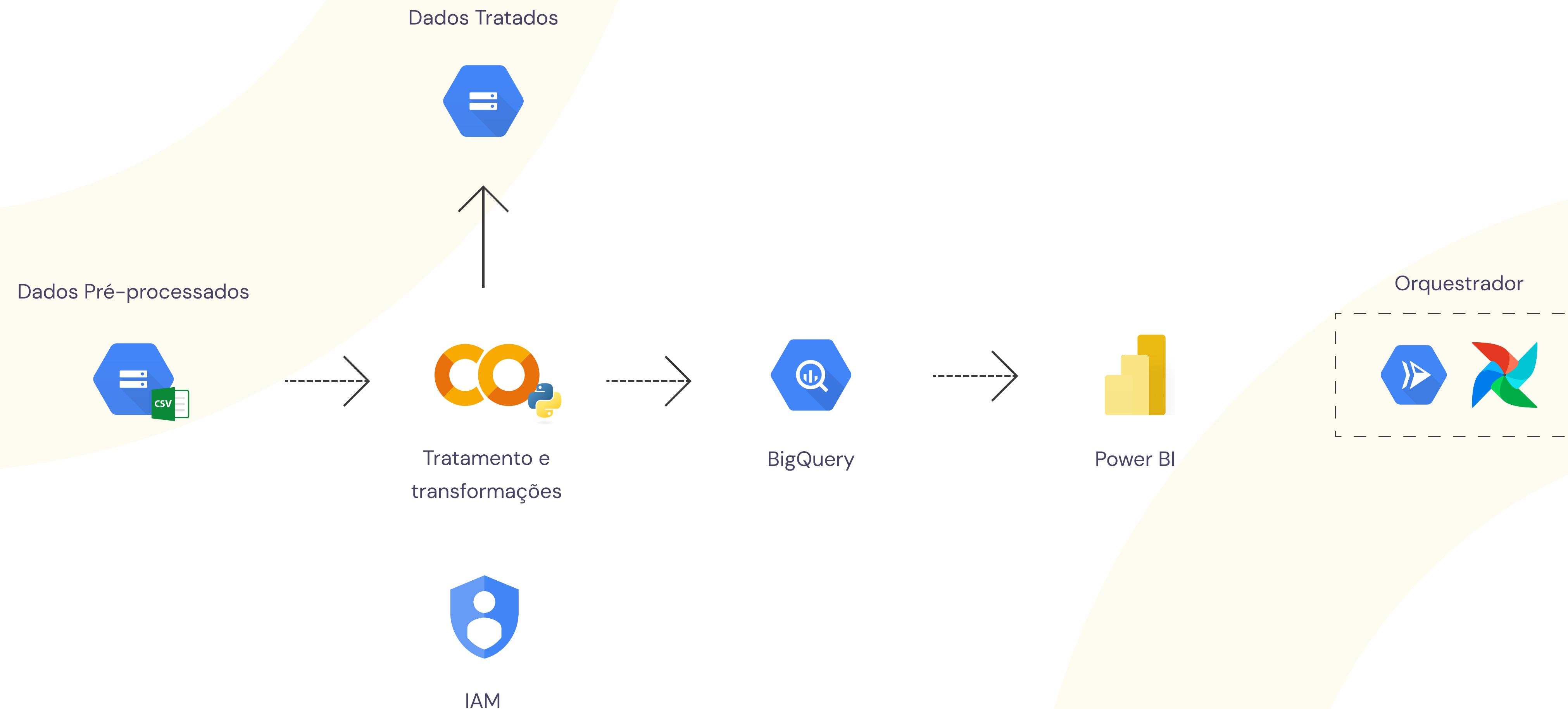


— A divisão para a conquista

Extração



Tratamento e Transformação



— O tratamento

Em imagens

Antes

mes_referencia	ano_referencia	cpf/id	nome	área	data_inicio	data_termino	duracao
novembro	2024	322-06-6396	Orson Rushsorth	atendimento	09/14/2024	9/24/2024	10.0
maio	2024	754-27-6437	Paxon Dudgeon	publicidade	9/30/2024	10/10/2024	10.0
janeiro	error	102-28-0836	Antoinette Mathias	produção	8/22/2024	9/1/2024	10.0
maio	2024	853-88-9859	Gav Sedger	produção	06/09/2024	6/19/2024	10.0

id_funcionario	nome	area	data_inicio	data_termino	duracao
322-06-6396	Orson Rushsorth	Atendimento	2024-09-14	2024-09-24	10
754-27-6437	Paxon Dudgeon	Publicidade	2024-09-30	2024-10-10	10
102-28-0836	Antoinette Mathias	Produção	2024-08-22	2024-09-01	10
853-88-9859	Gav Sedger	Produção	2024-06-09	2024-06-19	10

Depois



```
try:
    files.download("vagas_rs.csv")
    print("Download do arquivo CSV realizado com sucesso!\n")
except Exception as e:
    print(f"Erro ao fazer o download do arquivo CSV: {e}\n")

def write_csv_storage(arquivo):
    bucket_name = "projeto-dm9-final"
    blob_name = f"dados_tratados/{arquivo}.csv"
    storage_client = storage.Client()
    bucket = storage_client.bucket(bucket_name)
    blob = bucket.blob(blob_name)

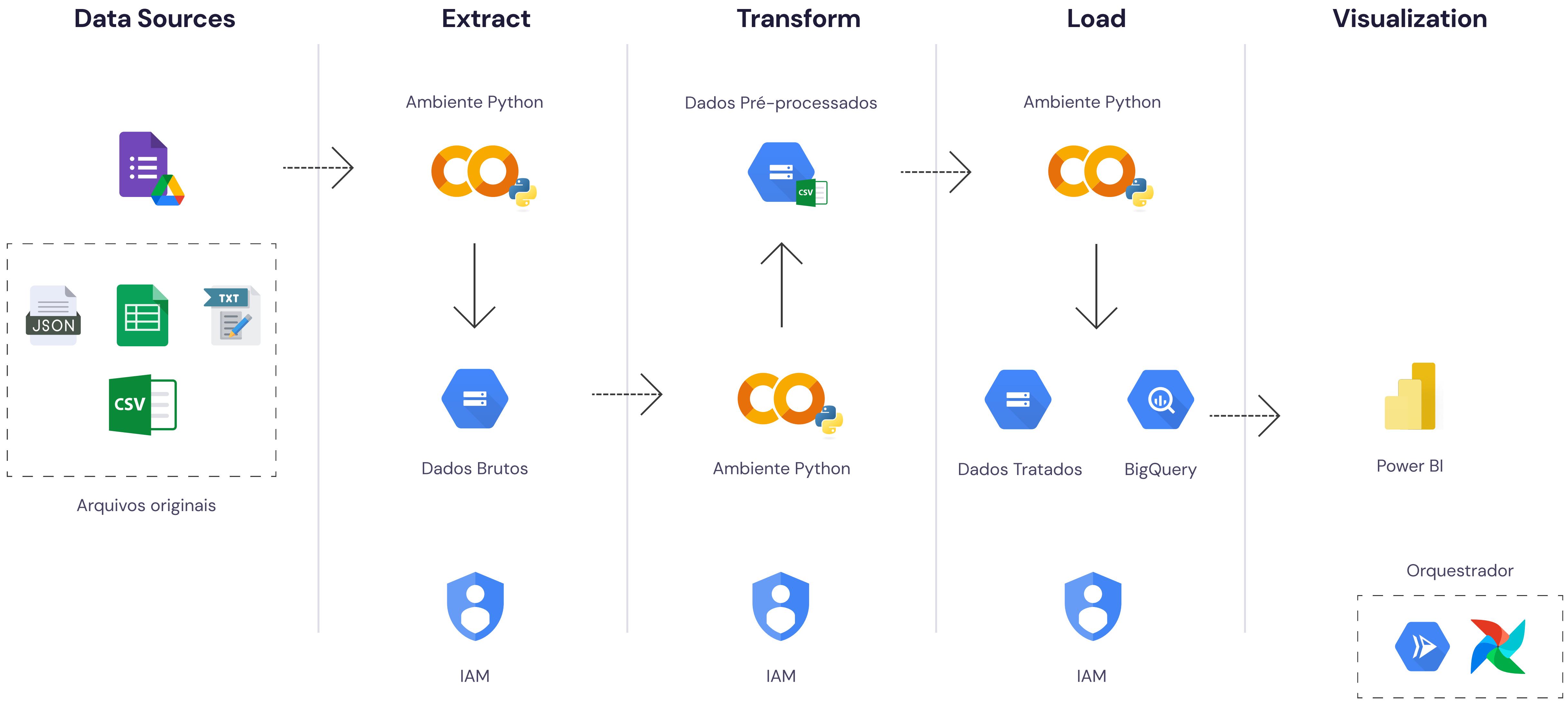
    blob.upload_from_filename(f"{arquivo}.csv")

write_csv_storage("vagas_rs")
print("Arquivo enviado com sucesso para a pasta 'dados_tratados'.\n")
table_id = "projeto-final-joshua.dm9.vagas_rs"

try:
    to_gbq(df_vagas_rs, table_id, project_id=project_id, if_exists="replace")
    print("\n\nEnvio para o BigQuery realizado com sucesso!")
except Exception as e:
    print(f"\n\nErro ao enviar os dados para o BigQuery: {e}")
```

O código







Refatoração e entrega



— A entrega



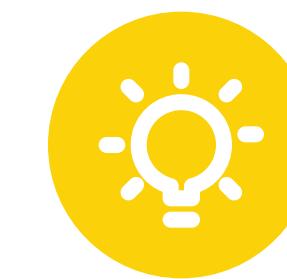
Refinamento do Código

Ajustamos scripts de ETL para melhorar eficiência e reduzir erros



Correção de Inconsistências

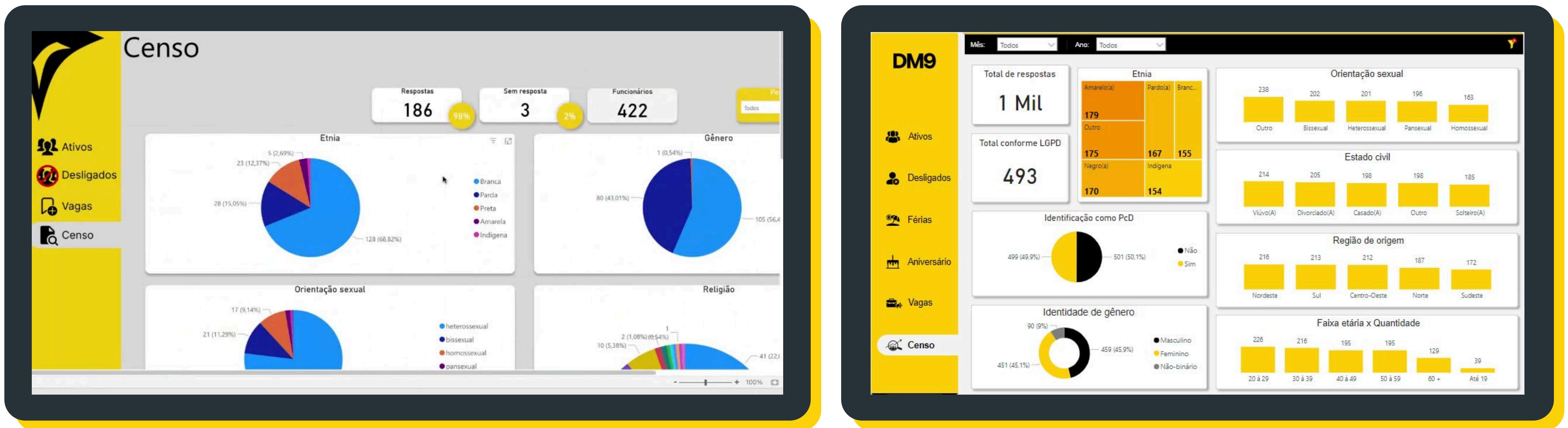
Identificamos e ajustamos dados inconsistentes que impactavam a qualidade dos dashboards



Qualidade dos Dados

Garantimos que as bases carregadas no BigQuery estavam corretas e bem estruturadas

Antes e depois



— Os gráficos

Pequenas mudanças

Grandes impactos

01 **Economia de tempo**

A automação eliminou tarefas manuais, reduzindo o tempo de processamento

02 **Confiabilidade dos dados**

A eliminação de inconsistências fortalece a precisão dos relatórios gerenciais

03 **Escalabilidade**

Possibilidade de integração com outras fontes, mantendo a flexibilidade do sistema

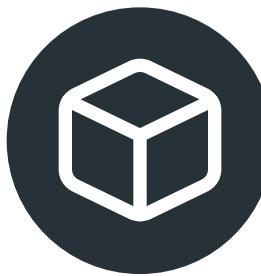
04 **Baixo custo**

Redução de custos operacionais ao eliminar retrabalho com planilhas manuais



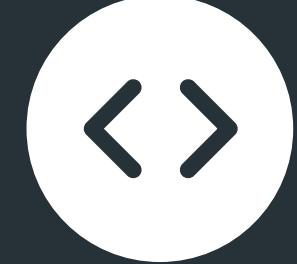
Desafios e dificuldades

O que enfrentamos durante o desenvolvimento do projeto
e como resolvemos



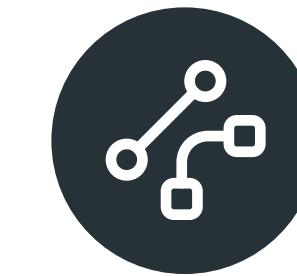
Limitações

Modelagem e análise de dados fictícios (simulados), visto que os dados reais não chegaram em tempo hábil



Planejamento

Elaboração de indicadores na fase final do projeto, antes da elaboração dos dashboards, gerou retrabalho. O Power BI não é a ferramenta mais indicada para tratamento de dados



Automação paliativa

Com base no estado atual da DM9 (receber planilhas e arquivos, tratá-las, enviar para a Big Query e elaborar dashboards no PBI ao invés de uma solução robusta de banco de dados



01

**Consumir dados
da fonte**

02

**Custo e
viabilidade**

03

**Normalizar as
tabelas e criar
atributos**

04

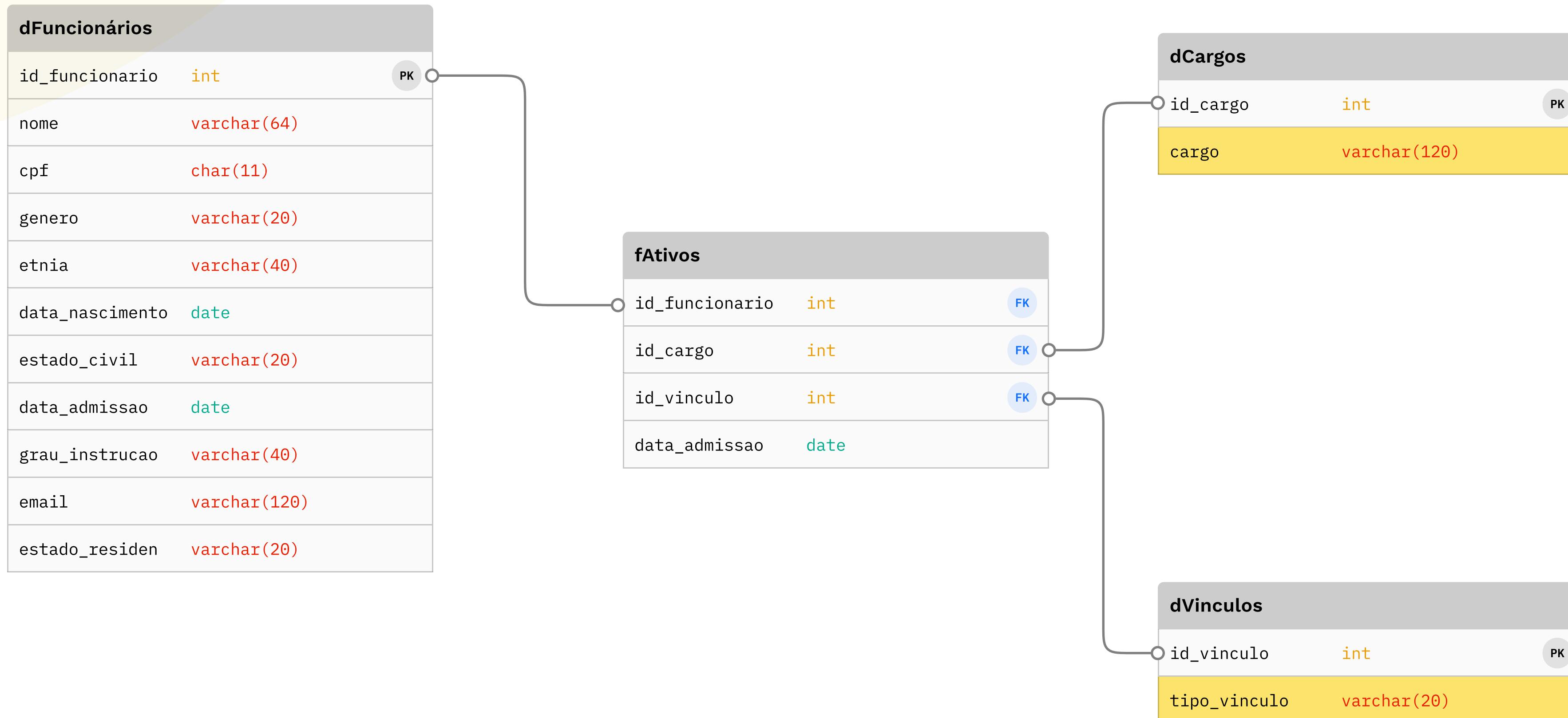
**Procedimentos de
ingestão
e atualização**

05

**Monitoramento e
escalabilidade**



Banco de Dados Relacional



```
Database changed
mysql> SHOW TABLES;
+-----+
| Tables_in_dpessoal |
+-----+
| dAgencias           |
| dAreas               |
| dCargos              |
| dFuncionarios        |
| dNiveis              |
| dVinculos            |
| fAtivos              |
| fFerias              |
| fPosicoes            |
+-----+
9 rows in set (0.04 sec)
```

```
+-----+-----+-----+-----+
| nome          | cargo          | tipo_vinculo | data_admissao |
+-----+-----+-----+-----+
| Pietro Pastor | Analista de Dados | Cooperado    | 2025-01-18   |
| Srta. Ísis Andrade | Social Media | Cooperado    | 2025-01-06   |
| Benicio Camargo | Gestor de Tráfego | Cooperado    | 2025-01-01   |
| Melissa Carvalho | Gerente de Projetos | CLT          | 2024-12-05   |
| Isadora Barbosa | Gerente de Projetos | Cooperado    | 2024-12-03   |
+-----+-----+-----+-----+
5 rows in set (0.04 sec)
```

```
+-----+-----+-----+-----+
| nome          | cargo          | tipo_vinculo | data_admissao |
+-----+-----+-----+-----+
| Pietro Pastor | Analista de Dados | PJ           | 2025-01-18   |
| Srta. Ísis Andrade | Social Media | PJ           | 2025-01-06   |
| Benicio Camargo | Gestor de Tráfego | PJ           | 2025-01-01   |
| Melissa Carvalho | Gerente de Projetos | CLT         | 2024-12-05   |
| Isadora Barbosa | Gerente de Projetos | PJ           | 2024-12-03   |
+-----+-----+-----+-----+
5 rows in set (0.03 sec)
```

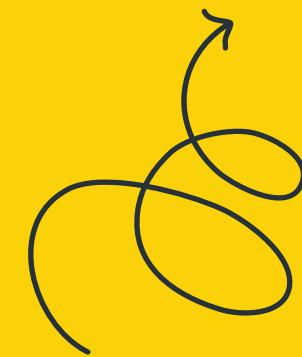
```
mysql> UPDATE dVinculos
      -> SET tipo_vinculo = 'Cooperado'
      -> WHERE id_vinculo = 2;
Query OK, 1 row affected (0.04 sec)
Rows matched: 1  Changed: 1  Warnings: 0
```





**Obrigado
pela atenção!**

Analista de Dados



Bombado!



Projeto SoulCode