

**UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE GESTÃO E NEGÓCIOS  
GRADUAÇÃO EM GESTÃO DA INFORMAÇÃO**

**MATHEUS CORRÊA GUIMARÃES**

**Implementação de BI: integração de dados do CRM ao power BI**

**ORIENTADOR: PROF. DR. JOSÉ EDUARDO FERREIRA LOPES**

**UBERLÂNDIA – MG**

**2024**

**MATHEUS CORRÊA GUIMARÃES**

**Implementação de BI: integração de dados do CRM ao Power BI**

Monografia apresentada ao Curso de Graduação em Gestão da Informação, da Universidade Federal de Uberlândia, como exigência parcial para a obtenção do título de Bacharel.

Orientador: Prof. Dr. José Eduardo Ferreira Lopes

**UBERLÂNDIA – MG**

**2024**

## RESUMO

Objetiva-se com este relato tecnológico descrever o processo de implantação de um ambiente de BI, com foco na integração de dados do sistema de CRM ao Power BI, para uma empresa do ramo do telemarketing, cujo objetivo era o de melhorar a qualidade e a rapidez na tomada de decisões. A ausência de um fluxo contínuo e automatizado de dados para suportar o processo de decisão em tempo hábil, devido aos complexos processos manuais, revelou a problematização que idealizou este projeto. Como solução, implantou-se um pipeline de dados dentro do ambiente Google Cloud Platform (GCP). O pipeline automatizou a extração, transformação e carregamento dos dados, permitindo que eles fossem atualizados automaticamente no Power BI. Como resultados alcançados, destaca-se a redução do tempo de processamento de dados, a melhoria na qualidade das informações disponíveis para as análises e a capacidade de tomada de decisão com base em dados próximo ao tempo real.

**Palavras-chave:** Business Intelligence; Google Cloud Platform; ETL; Data Warehouse; Pipeline de dados.

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>5</b>
<b>2 FUNDAMENTAÇÃO TEÓRICA .....</b>	<b>5</b>
<b>2.1 O QUE É UM DATA WAREHOUSE .....</b>	<b>6</b>
<b>2.2 MODELOS DE DATA WAREHOUSE: ABORDAGENS DE INMON E KIMBALL .....</b>	<b>6</b>
<b>3 CONTEXTO INVESTIGADO E SITUAÇÃO PROBLEMA.....</b>	<b>8</b>
<b>4 INTERVENÇÃO ADOTADA.....</b>	<b>9</b>
<b>5 RESULTADOS ALCANÇADOS (OU ESPERADOS) .....</b>	<b>11</b>
<b>6 CONSIDERAÇÕES FINAIS.....</b>	<b>13</b>
<b>7 REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>14</b>

## 1 INTRODUÇÃO

O avanço tecnológico nas últimas décadas proporcionou às empresas novas formas de coletar, processar e analisar dados, transformando-os em ativos estratégicos fundamentais para a tomada de decisões. No setor de telemarketing, onde a competição é acirrada e as margens de lucro são estreitas, a utilização eficiente dos dados pode significar a diferença entre o sucesso e o fracasso de uma empresa. Neste contexto, o *Business Intelligence* (BI) se apresenta como uma ferramenta indispensável, capaz de transformar grandes volumes de dados em informações úteis para tomada de decisão. (Sharda; Delen; Turban, 2020, p. 14)

Um dos principais problemas identificados na empresa de telemarketing foi a ausência de um fluxo contínuo e automatizado de dados que pudesse suportar o processo de decisão em tempo hábil. A dependência de processos manuais para a consolidação dos dados do sistema de *Customer Relationship Management* (CRM) dificultava a obtenção de informações precisas e atualizadas, o que comprometia a eficácia das análises realizadas, inviabilizava o desenvolvimento dos indicadores de KPIs definidos e, conseqüentemente, a qualidade das decisões tomadas pela empresa.

Para superar esse desafio, foi necessário desenvolver uma solução tecnológica capaz de integrar os dados do CRM ao Power BI de forma automatizada e próxima ao tempo real. A implementação desse pipeline de dados além de resolver o problema da falta de automatização, também melhoraria significativamente a qualidade das informações disponibilizadas para análise.

Nesse contexto, objetiva-se com este relato tecnológico descrever o processo de implantação de um ambiente de BI, com foco na integração de dados do sistema de CRM ao Power BI, para uma empresa no ramo do telemarketing, a fim de melhorar a qualidade e a rapidez na tomada de decisões.

## 2 FUNDAMENTAÇÃO TEÓRICA

Ao longo dos anos, com o crescente volume de dados gerados, e o forte impacto que eles têm apresentado nas tomadas de decisões, o *Business Intelligence* vem sendo cada vez mais difundido dentro do meio empresarial. O termo BI foi popularizado por Dresner (1989), que definiu o tema como uma junção de conceitos, metodologias e ferramentas que ajudam as organizações a tomar decisões baseadas em dados históricos.

Segundo Kimball (1996), o principal objetivo do *Business Intelligence* é permitir com que os dados contidos em *data warehouses* sejam apresentados de maneira compreensível e acessível para gestores, de modo a suportar decisões rápidas e embasadas.

## 2.1 O que é um data warehouse

A construção de um *data warehouse* é uma das principais etapas para a criação de uma solução de BI, onde será permitido com que todas as informações sejam acessadas de forma clara e coesa.

Para Inmon (2005) um *data warehouse* é uma coleção de dados orientados por assunto, integrados, variáveis com o tempo e não voláteis, que suportam a tomada de decisão. Inmon defende uma construção de um modelo centralizado, com os dados sendo organizados em uma estrutura normalizada, garantindo maior consistência e integridade.

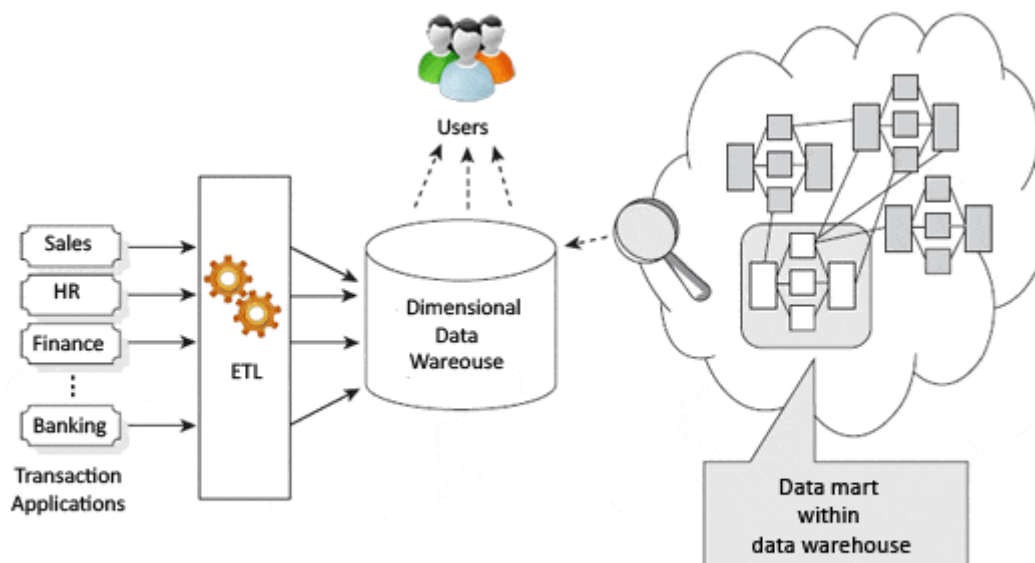
Em contrapartida, segundo Kimball (1996), um *data warehouse* é um sistema que armazena dados organizados de forma a facilitar a criação de relatórios e análises, utilizando um modelo dimensional em uma arquitetura desnormalizada, e de baixo para cima, com foco na otimização de consultas que facilite o processo de extração de informações úteis para o negócio.

## 2.2 Modelos de Data Warehouse: Abordagens de Inmon e Kimball

Atualmente, existem duas principais abordagens de *data warehouse* que são amplamente implementadas; o modelo "*Top-Down*" de William H. Inmon e o modelo "*Bottom-Up*" de Ralph Kimball.

Kimball (1996) define o *data warehouse* como um ambiente de dados orientado ao usuário final, focado em facilitar e otimizar o acesso às informações para análise e suporte à tomada de decisões. Diante disso, Kimball defende um *data warehouse* como uma arquitetura descentralizada de baixo para cima (*Bottom-UP*), com o foco na segmentação por áreas de negócio através de *data marts*, focando em uma maior eficiência ao acesso de informações.

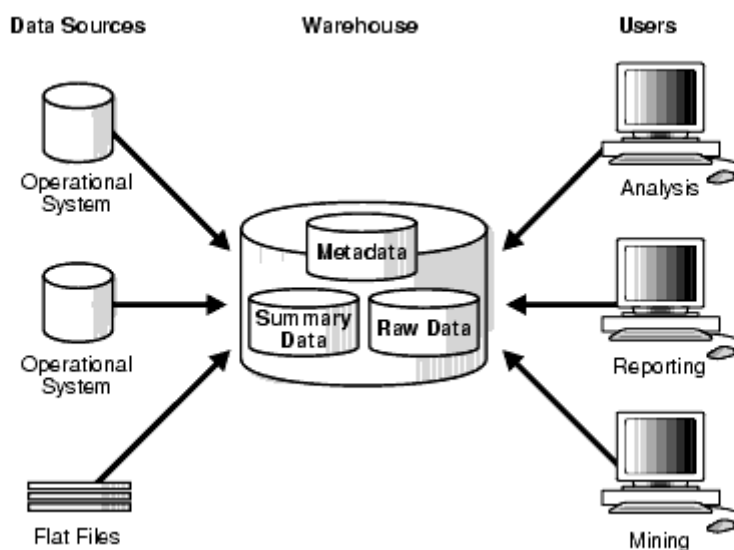
Figura 1 - Arquitetura do Data Warehouse por Ralph Kimball



Fonte: Zenut (2024)

Em contrapartida, Inmon (1992) destaca um *data Warehouse* como um ambiente de dados mais robusto centralizado, em um formato *top-down* (Cima para baixo). Nesta abordagem, todos os dados da organização são integrados em um único repositório, estruturado de forma normalizada para garantir a consistência, integridade e governança das informações, a partir deste repositório central, *data marts* podem ser criados para atender a necessidades específicas de análise. Enquanto Kimball propôs um modelo com foco na eficiência e orientado ao usuário, Inmon sugere uma arquitetura que prioriza a integridade e consistência dos dados.

Figura 2 - Data Warehouse por William H. Inmon



Fonte: Atera (2024)

### 3 CONTEXTO INVESTIGADO E SITUAÇÃO PROBLEMA

A empresa do setor de telemarketing, fundada em 2022, conta atualmente com cerca de 50 funcionários e tem como foco oferecer serviços e soluções empresariais personalizadas no atendimento ao cliente, conta com centenas ou até mesmo milhares de registros gerados diariamente, que se analisados corretamente, geram um grande valor organizacional.

Por se tratar de uma empresa de telemarketing com o foco no atendimento ao cliente, é necessário o estabelecimento de metas claras, tanto para a equipe quanto para a gestão, a fim de analisar fatores como; rendimento do time, qualidade dos atendimentos ao cliente, quantidade de negociações realizadas, entre outros indicadores que para serem eficazes, dependem de dados consistentes para análise.

Diante desta situação, encontrou-se um grande problema no processo analítico. Embora as métricas, indicadores e as visualizações de dados já estivessem mapeadas, era ainda necessário realizar o processo de coleta, transformação e carregamento de dados manualmente e diariamente. Uma vez que nem todos os colaboradores possuíam tais conhecimentos técnicos, o processo se limitava apenas a poucos funcionários, demandando um tempo que na maioria das vezes não era possível ser disponibilizado. Além disso, a empresa não possuía um repositório de dados seguro para o armazenamento



dos registros, o que impedia o acesso a um histórico dos dados e tão pouco, conseguiam garantir que o tratamento dos dados era realizado de maneira correta, uma vez que o processo manual se torna mais propício a falhas. Diante disso, a ausência de um *data warehouse*, o grande tempo gasto com o processo manual, e a má qualidade dos dados inviabilizava todo o acompanhamento de análise e estipulação de metas da empresa.

#### **4 INTERVENÇÃO ADOTADA**

Devido aos fatores apresentados anteriormente, que dificultavam a qualidade dos dados, inviabilizando a análise e o desenvolvimento dos indicadores já mapeados, foi identificada a oportunidade de melhoria nas etapas de extrair, transformar, carregar e analisar os dados, substituindo todo o processo manual por um totalmente automatizado, além da construção de um *data warehouse*, conseqüentemente, melhorando a precisão e a eficiência das análises de dados.

Durante o planejamento da solução, definiu-se como requisitos de desenvolvimento, algo que suprisse as necessidades de negócio, automatizado, com a mínima necessidade de trabalho humano no dia a dia, em um ambiente que não fosse necessário depender de máquinas e servidores locais ligados para a execução dos procedimentos. O *pipeline* deveria ser escalável a fim de suportar grandes volumes de dados futuramente, disponibilizar um repositório central de dados e além de tudo, com um baixo custo financeiro.

Pensando em todas as necessidades pontuadas acima, acordou-se com os *stakeholders* que uma solução em nuvem seria a ideal, uma vez que desta forma supririam as necessidades de independência de servidores locais, além de permitir a construção de um *data Warehouse* escalável de acordo com a necessidade do cliente. Deste modo, estudou-se as seguintes plataformas de computação em nuvem: *Amazon Web Service (AWS)*, *Microsoft Azure* e *Google Cloud Platform (GCP)*.

Analisando a fundo cada plataforma, concluiu-se que a *Amazon Web Services* seria uma ótima opção em casos de mais controle e de uma infraestrutura mais flexível, mas poderia ser um pouco complexa para um pipeline menor. Enquanto a plataforma *Microsoft Azure* seria ideal em um cenário em que será utilizado produtos da Microsoft e busca integração direta com essas ferramentas, mas é em geral, mais caro em comparação com GCP para projetos menores.

Portanto, analisando qual plataforma atendia melhor os requisitos, chegou-se à conclusão de que o *Google Cloud Platform* (GCP) é, em geral, a melhor escolha para construir um pipeline pequeno, devido ao modelo de preço simples e sob demanda, uma interface mais fácil de utilizar, e a camada gratuita. Além do *BigQuery* como *data Warehouse* ser uma excelente escolha para repositório de dados, e a escalabilidade automática do GCP facilita a adaptação futura.

A solução adotada envolveu a construção de um pipeline de dados automatizado dentro do ambiente *Google Cloud Platform*. A extração dos dados foi realizada através da biblioteca *requests* desenvolvida através da linguagem de programação Python, permitindo o consumo através da API fornecida pela empresa responsável pelo *Customer Relationship Management*. O tratamento de dados dentro das *Google Cloud Functions* foi feito utilizando a biblioteca *Pandas*, garantindo que os dados estivessem limpos e prontos para serem carregados no *Google BigQuery* utilizado como *data Warehouse* da empresa. Parte deste *script* pode ser visualizado na Figura 3.

Figura 3 - Script Extração Dentro da Cloud Function

```
import functions_framework
import pandas as pd
import requests
import json
from io import StringIO
from datetime import datetime, timedelta
import pytz
from google.cloud import bigquery

def extract(url, tokenEstrutura, tokenUsuario, painelId, formato):
    try:
        fuso_horario = pytz.timezone('America/Sao_Paulo')
        atual = datetime.now(fuso_horario)
        anterior = atual - timedelta(hours=1, minutes=30)
        data_atual = atual.strftime("%Y-%m-%d %H:%M:%S")
        data_anterior = anterior.strftime("%Y-%m-%d %H:%M:%S")

        payload = {
            "tokenEstrutura": tokenEstrutura,
            "tokenUsuario": tokenUsuario,
            "dataHoraInicioCarga": data_anterior,
            "dataHoraFimCarga": data_atual,
            "painelId": painelId,
            "outputFormat": formato
        }
        headers = {
            'Content-Type': 'text/plain',
            'Cookie': '__cf1b=02DiuHcRebXBbQZs3gX28EM2MeLsdaT3jc2MMtm36LJzp'
        }

        response = requests.post(url, headers=headers, data=json.dumps(payload), ensure_ascii=False)
        response.raise_for_status() # Levanta uma exceção para códigos de status de erro

        csv_data = response.text
        df = pd.read_csv(StringIO(csv_data), delimiter=',')
        return df

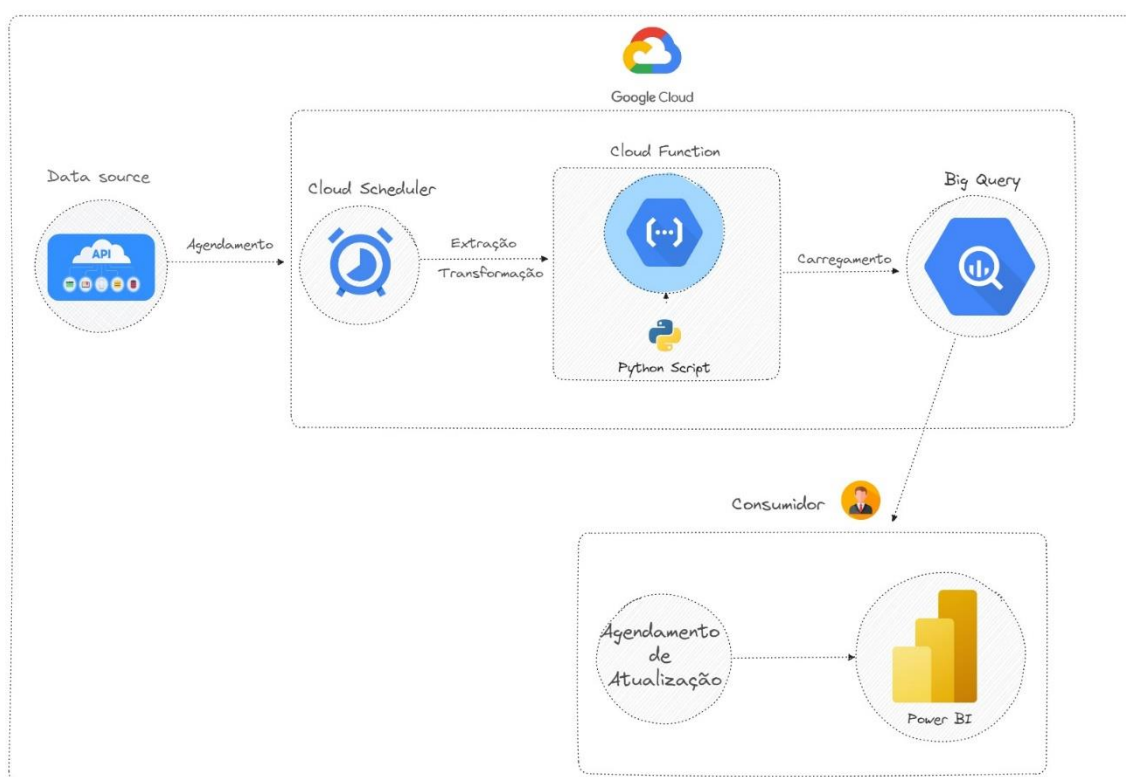
    except requests.RequestException as e:
        print(f"Erro na requisição: {e}")
        return None
```

Fonte: Autoria própria

Para garantir a execução contínua, todo o processo foi orquestrado através de uma rotina agendada pelo *Google Cloud Scheduler*, de modo com que as funções no *Google Cloud Function* fossem executadas periodicamente em um intervalo de 90 minutos, carregando os dados para o *BigQuery* de forma incremental.

Por fim, a integração do *Data Warehouse* com o Power BI foi feita através da conexão nativa do Power BI e configurado uma rotina de atualização periódica, assegurando que as visualizações de dados e relatórios estejam constantemente atualizadas, viabilizando análises em tempo real e suporte à tomada de decisões estratégicas pela empresa.

Figura 4 - Fluxograma Pipeline de Dados



Fonte: Autoria própria

## 5 RESULTADOS ALCANÇADOS (OU ESPERADOS)

Após a implementação do pipeline, cada fluxo de trabalho foi executado conforme o esperado, permitindo a automatização de todo o processo. A companhia obteve uma série de melhorias significativas em seus processos analíticos. Anteriormente, a empresa enfrentava grandes desafios na coleta, transformação e análise dos dados devido à necessidade de realizar essas tarefas manualmente, o que consumia tempo e comprometia

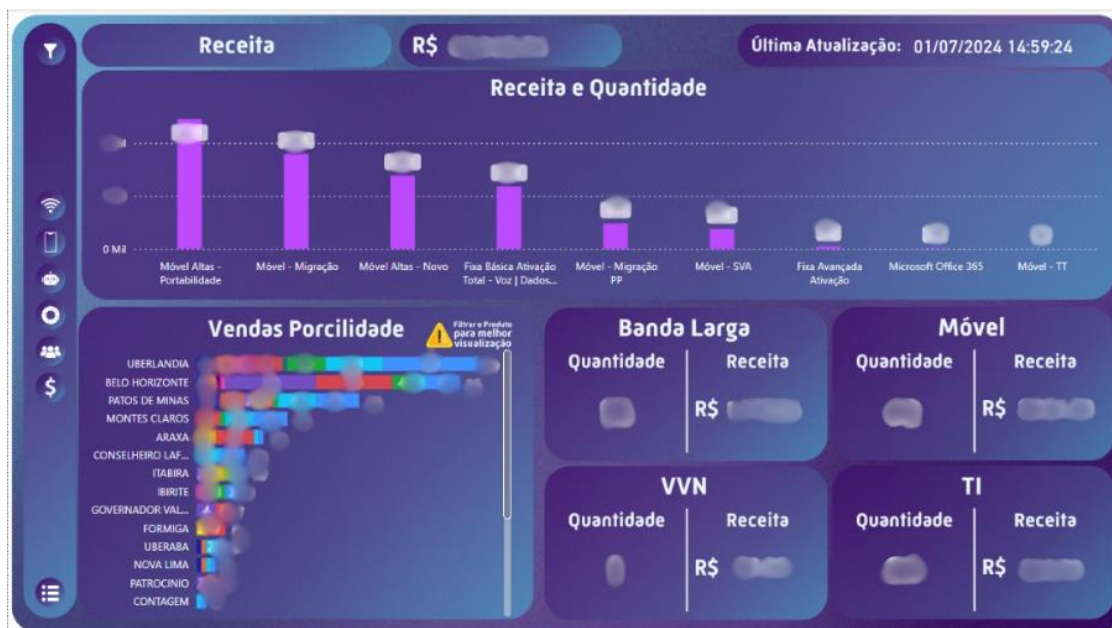
a qualidade das informações. Após a implementação do serviço, os processos manuais foram descartados e substituídos pela automatização e integração dos dados.

Dentre os benefícios alcançados, temos como principal ganho, a redução no tempo de processamento dos dados. A automatização do pipeline eliminou a necessidade de intervenções manuais, por consequência, reduziu significativamente o tempo necessário para a coleta e transformação dos dados. Antes, o processo manual consumia várias horas por dia, enquanto o pipeline automatizado processa os dados em intervalos de 90 minutos, sem necessidade de trabalhos humanos. Ademais, obteve um ganho na qualidade dos dados com o uso de um *Data Warehouse* centralizado no *Google BigQuery*, garantindo consistência, integridade, histórico e escalabilidade dos dados. O processo automatizado de ETL (Extração, Transformação e Carregamento) reduziu os erros humanos e aumentou a qualidade das informações analisadas.

Um dos aspectos fundamentais em uma empresa em crescimento, voltada para o ramo do telemarketing, é a competência de tomada de decisão mais ágil, e a integração do *pipeline* ao *Power BI* permitiu que os gestores *startup* tivessem acesso em tempo real aos relatórios e *dashboards*, permitindo análises sempre atualizadas. Aumentando assim, a capacidade de responder rapidamente a desafios operacionais, ajustar metas, e acima de tudo, tomar decisão.

A Figura 5 ilustra um dos painéis que dão suporte ao processo de tomada de decisão.

Figura 5 - Dashboard



Fonte: Autoria própria

## 6 CONSIDERAÇÕES FINAIS

Objetivou-se com este relato tecnológico descrever o processo de implantação de um ambiente de BI, com foco na integração de dados do sistema de CRM ao Power BI, para uma empresa no ramo do telemarketing, a fim de melhorar a qualidade e a rapidez na tomada de decisões. A ausência de um fluxo contínuo e automatizado de dados para suportar o processo de decisão em tempo hábil, devido os complexos processos manuais, revelou a problematização que idealizou este projeto. Como solução, implantou-se um pipeline de dados dentro do ambiente Google Cloud Platform (GCP). O pipeline automatizou a extração, transformação e carregamento dos dados, permitindo que eles fossem atualizados automaticamente no Power BI. Como resultados alcançados, destaca-se a redução do tempo de processamento de dados, a melhoria na qualidade das informações disponíveis para as análises e a capacidade de tomada de decisão com base em dados próximo ao tempo real.

Adicionalmente, destaca-se o fato de este profissional, autor deste relato, ter participado do projeto de forma ativa, permitindo consolidar os aprendizados iniciados durante a graduação.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

ASTERA. **Conceitos de data warehouse: abordagem Kimball vs. Inmon** <<https://www.astera.com/pt/type/blog/data-warehouse-concepts/>> Acesso em: 20 de set. 2024.

KIMBALL, Ralph. **The data warehouse toolkit: practical techniques for building dimensional data warehouses**. John Wiley & Sons, Inc., 1996.

PANIAN, Zeljko. **The evolution of business intelligence: from historical data mining to mobile and location-based intelligence**. In: WSEAS International Conference on Recent Researches in Business and Economics, ISBN. 2012.

SHARDA, Ramesh; DELEN, Dursun; TURBAN, Efraim. **Business Intelligence e Análise de Dados para Gestão do Negócio**. 4. ed. São Paulo: Bookman, 2019.

INMON, William H. **Building the data warehouse**. John wiley & sons, 2005.

ZENUT. **Ralph Kimball Data Warehouse Architecture** <<https://www.zentut.com/data-warehouse/ralph-kimball-data-warehouse-architecture/>> Acesso em: 20 de set. 2024.