



UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO TECNOLÓGICO  
DEPARTAMENTO DE AUTOMAÇÃO E SISTEMAS  
CURSO DE GRADUAÇÃO EM ENGENHARIA DE CONTROLE E AUTOMAÇÃO

Alex Lopes

**Desenvolvimento de uma Arquitetura de Dados Otimizada  
para Análise de Inteligência de Negócio**

Florianópolis  
2024

Alex Lopes

**Desenvolvimento de uma Arquitetura de Dados Otimizada  
para Análise de Inteligência de Negócio**

Relatório final da disciplina DAS5511 (Projeto de Fim de Curso) como Trabalho de Conclusão do Curso de Graduação em Engenharia de Controle e Automação da Universidade Federal de Santa Catarina em Florianópolis.

Orientador: Prof. Daniel Coutinho, Dr.

Supervisor: Germain Velard, Eng.

Florianópolis

2024

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.  
Dados inseridos pelo próprio autor.

Lopes, Alex Inocencio

Desenvolvimento de uma arquitetura de dados otimizada  
para análise de inteligência de negócio / Alex Inocencio  
Lopes ; orientador, Daniel Ferreira Coutinho, 2024.

99 p.

Trabalho de Conclusão de Curso (graduação) -  
Universidade Federal de Santa Catarina, Centro Tecnológico,  
Graduação em Engenharia de Controle e Automação,  
Florianópolis, 2024.

Inclui referências.

1. Engenharia de Controle e Automação. 2. Dados. 3.  
Inteligência de negócio. 4. Financeiro. 5. Energia. I.  
Coutinho, Daniel Ferreira. II. Universidade Federal de  
Santa Catarina. Graduação em Engenharia de Controle e  
Automação. III. Título.

Alex Lopes

**Desenvolvimento de uma Arquitetura de Dados Otimizada  
para Análise de Inteligência de Negócio**

Esta monografia foi julgada no contexto da disciplina DAS5511 (Projeto de Fim de Curso) e aprovada em sua forma final pelo Curso de Graduação em Engenharia de Controle e Automação

Florianópolis, 20 de maio de 2024.

Prof. Hector Bessa Silveira, Dr.  
Coordenador do Curso

**Banca Examinadora:**



Documento assinado digitalmente

**Daniel Ferreira Coutinho**

Data: 02/08/2024 16:05:24-0300

CPF: \*\*\*.169.980-\*\*

Verifique as assinaturas em <https://v.ufsc.br>

Prof. Daniel Coutinho, Dr.

Orientador  
UFSC/CTC/DAS

Germain Velard, Eng.  
Supervisor  
Engie Brasil Participações Ltda.



Prof. Ricardo Rabelo, Dr.

Avaliador  
UFSC/CTC/DAS

Prof. Eduardo Camponogara, Dr.  
Presidente da Banca  
UFSC/CTC/DAS

Este trabalho é dedicado aos meus colegas de classe e  
aos meus queridos pais.

## **AGRADECIMENTOS**

Gostaria de agradecer meus pais e minha irmã que me proporcionaram a oportunidade de estudar na Universidade Federal de Santa Catarina, não apenas financeiramente e apoiando nas minhas decisões durante a graduação, mas também previamente moldando e me preparando para os desafios futuros. Ao professor Daniel Coutinho por aceitar ser meu orientador, e por ser um grande incentivador dos seus alunos à terem foco em excelência e no aprendizado constante. Ao meu chefe Germain Velard e meus colegas pela confiança ao me incorporar no time de controle de riscos da Engie Brasil, por toda atenção e compartilhamento de conhecimento. Com certeza eles e meus demais colegas tornaram o ambiente muito mais estimulante e agradável, facilitando meu crescimento pessoal e profissional ao longo dessa jornada. Aos meus amigos da graduação e em especial ao Alcioni Panatto, por serem uma rede de apoio que me deram a base para superar desafios e seguir a diante mesmo nos momentos mais difíceis. E, por último, ao Departamento de Automação e Sistemas e à Universidade Federal de Santa Catarina por proporcionarem uma educação superior de excelência.

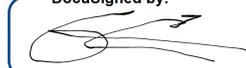
*“Inteligência é a capacidade de absorver informação em tempo real.  
De fazer perguntas que façam sentido. É ter boa memória.  
É traçar pontes entre assuntos que não  
parecem estar relacionados e inovar ao fazer essas conexões.”  
(GATES, 2020)*

## DECLARAÇÃO DE PUBLICIDADE

Florianópolis, 20 de Junho de 2024.

Sigilo Na condição de representante da Engie Brasil Participações Ltda. na qual o presente trabalho foi realizado, declaro não haver ressalvas quanto ao aspecto de sigilo ou propriedade intelectual sobre as informações contidas neste documento, que impeçam a sua publicação por parte da Universidade Federal de Santa Catarina (UFSC) para acesso pelo público em geral, incluindo a sua disponibilização *online* no Repositório Institucional da Biblioteca Universitária da UFSC. Além disso, declaro ciência de que o autor, na condição de estudante da UFSC, é obrigado a depositar este documento, por se tratar de um Trabalho de Conclusão de Curso, no referido Repositório Institucional, em atendimento à Resolução Normativa n° 126/2019/CUn.

Por estar de acordo com esses termos, subscrevo-me abaixo.

DocuSigned by:  
  
4CD72D1B29B4401...

---

Germain Velard  
Engie Brasil Participações Ltda.

## RESUMO

Este projeto de fim de curso consiste no desenvolvimento de uma arquitetura de dados empresariais que armazena informações em versões e os transforma para realizar análises estratégicas, mostrar indicadores e métricas importantes de desempenho, e prover informações gráficas através de painéis interativos (conhecidos no meio empresarial como relatórios). Atualmente, a área de financeira de controle de risco da empresa Engie Brasil Participações possui uma arquitetura de dados não estruturada, sendo desta forma suscetível a erros e falhas os quais podem passar despercebidos, ocasionando tomadas de decisões baseadas em dados não processados corretamente ou até mesmo incorretos. Além disso, a interação com usuários (e.g., avaliadores de desempenho e estrategistas) não é muito amigável, de modo que outros colaboradores têm dificuldade de utilizar a ferramenta final de visualização. Os resultados obtidos neste projeto levaram a construção de uma ferramenta de processamento e estruturação do banco de dados da empresa, podendo ser resumido em dois blocos: (i) o bloco de extração, processamento e transformação de dados de diferentes fontes em diferentes domínios; e (ii) bloco de visualização e avaliação onde os dados são analisados com auxílio de uma ferramenta iterativa.

**Palavras-chave:** Inteligência de negócio. Dados. Energia. Financeiro. Indicadores de desempenho

## ABSTRACT

This final project involves the development of an enterprise data architecture that stores information in versions and transforms it for strategic analysis, displaying key performance indicators and metrics, and providing graphical information through interactive dashboards. Currently, the financial risk control department at Engie Brasil Participações has an unstructured data architecture, making it susceptible to unnoticed errors and failures, which can lead to decision-making based on improperly processed or even incorrect data. Moreover, the user interaction (e.g., performance evaluators and strategists) is not very user-friendly, resulting in other employees finding it difficult to use the final visualization tool. The results obtained from this project led to the construction of a tool for processing and structuring the company's database, which can be summarized into two blocks: *(i)* the data extraction, processing, and transformation block from different sources across different domains; and *(ii)* the visualization and evaluation block where the data is analyzed with the aid of an interactive tool.

**Keywords:** Business intelligence. Data. Energy. Financial. Key performance indicators.

## LISTA DE FIGURAS

Figura 1 – Visão geral da ajuda do CRISP-DM. . . . .	18
Figura 2 – Logo da empresa Engie Brasil. . . . .	21
Figura 3 – Painel de dados da ENGIE Brasil. . . . .	21
Figura 4 – ENGIE no Brasil. . . . .	22
Figura 5 – Visão Geral da Cadeia setor energético. . . . .	23
Figura 6 – Matriz Elétrica Brasileira em 2022. . . . .	24
Figura 7 – Painel de Preço de Liquidação das Diferenças no ambiente da CCEE. . . . .	25
Figura 8 – Diagrama organizacional ENGIE Brasil. . . . .	26
Figura 9 – Visão dos dados de energia. . . . .	27
Figura 10 – Modelo de Banco de dados. . . . .	29
Figura 11 – Modelo genérico banco de dados relacional. . . . .	31
Figura 12 – Atributos e Registros de banco de dados relacionais. . . . .	31
Figura 13 – Tipos de relacionamentos de banco de dados. . . . .	32
Figura 14 – Esquema Nuvem da Microsoft. . . . .	33
Figura 15 – Esquema gateway. . . . .	34
Figura 16 – Data Warehouse genérico. . . . .	35
Figura 17 – Esquema estrutura Data Marts. . . . .	37
Figura 18 – Tabela Funções Data Mart. . . . .	38
Figura 19 – Esquema Processo ETL. . . . .	39
Figura 20 – Interface da ferramenta Power Query. . . . .	40
Figura 21 – Dado x Informação x Conhecimento. . . . .	41
Figura 22 – Interface Power BI. . . . .	42
Figura 23 – Mapeamento processo BPMN. . . . .	46
Figura 24 – Visão Data Warehouse Engie Brasil Energia. . . . .	47
Figura 25 – Pasta do Sharepoint. . . . .	47
Figura 26 – Diagrama UML Casos de Uso da proposta de solução do projeto. . . . .	51
Figura 27 – Diagrama UML de Sequência da proposta de solução do projeto. . . . .	52
Figura 28 – Construção Data Mart. . . . .	54
Figura 29 – Esquema Entendimento dos dados. . . . .	55
Figura 30 – Tabela de dados Data Mart do Balanço EBE. . . . .	56
Figura 31 – Tabela de dados da plataforma BBCE. . . . .	57
Figura 32 – Pasta Sharepoint. . . . .	58
Figura 33 – Fluxo de dados. . . . .	59
Figura 34 – Fluxo do ETL na ferramenta Power Query Online. . . . .	60
Figura 35 – Painel inicial da extração de dados. . . . .	60
Figura 36 – Extração pasta Sharepoint. . . . .	61
Figura 37 – Extração DW Amazon Redshift. . . . .	61

Figura 38 – Interface Power Query Online. . . . .	62
Figura 39 – Transformação de dados da tabela de versões do Balanço Unificado. . . . .	63
Figura 40 – União de tabelas (Merge). . . . .	64
Figura 41 – Tabela final Balanço Unificado. . . . .	65
Figura 42 – Atributo Produto. . . . .	66
Figura 43 – Etapas da transformação de dados da BBCE. . . . .	66
Figura 44 – Etapas de transformação dCalendar. . . . .	67
Figura 45 – Código DAX: número de dias operados na BBCE. . . . .	68
Figura 46 – Código DAX: liquidez do produto do próximo ano. . . . .	68
Figura 47 – Código DAX: Nível de contratação. . . . .	69
Figura 48 – Configuração de agendamento da atualização. . . . .	70
Figura 49 – Relacionamentos Data Mart Controle de Risco. . . . .	71
Figura 50 – Página Portfólio do relatório de Controle de Risco. . . . .	72
Figura 51 – Página Liquidez na BBCE do relatório de Controle de Risco. . . . .	73
Figura 52 – Página Nível de Contratação do relatório de Controle de Risco. . . . .	74
Figura 53 – Página Preços do relatório de Controle de Risco. . . . .	75
Figura 54 – Tabela resultado formulário de feedback. . . . .	78

## LISTA DE ABREVIATURAS E SIGLAS

ETL	<i>Extract, Transform and Load</i>
DW	<i>Data Warehouse</i>
DM	<i>Data Mart</i>
BBCE	Balcão Brasileiro de Comercialização de Energia
CRISP DM	<i>Cross Industry Standard Process for Data Mining</i>
CCEE	Câmara de Comercialização de Energia Elétrica
DBMS	<i>Data Base Management System</i>
EBE	Engie Brasil Energia
EBP	Engie Brasil Participações
ACL	<i>Ambiente de Contratação Livre</i>
ACR	<i>Ambiente de Contratação Regulado</i>
KNN	<i>K-nearest neighbor</i>
BPMN	<i>Business Process Model and Notation</i>
ANEEL	Agência Nacional de Energia Elétrica
BI	<i>Business Intelligence</i>
PLD	Preço de Liquidação das Diferenças
ONS	Operador Nacional do Sistema Elétrico
SIN	Sistema Interligado Nacional
CSV	<i>Comma Separated Values</i>
SQL	<i>Structured Query Language</i>
SE	Sudeste
TI	Tecnologia da Informação
UML	Linguagem de Modelagem Unificada

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>15</b>
1.1	OBJETIVOS	16
<b>1.1.1</b>	<b>Objetivo Geral</b>	<b>16</b>
<b>1.1.2</b>	<b>Objetivos Específicos</b>	<b>16</b>
1.2	METODOLOGIA	17
1.3	ESTRUTURA DO DOCUMENTO	20
<b>2</b>	<b>EMPRESA ENGIE BRASIL PARTICIPAÇÕES</b>	<b>21</b>
<b>2.0.1</b>	<b>Setor de energia</b>	<b>22</b>
<b>2.0.2</b>	<b>Área financeira de controle de riscos</b>	<b>25</b>
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>29</b>
3.1	BANCO DE DADOS	29
<b>3.1.1</b>	<b>Banco de dados Relacional</b>	<b>30</b>
3.2	TECNOLOGIA NA NUVEM	32
<b>3.2.1</b>	<b>Gateway de dados</b>	<b>34</b>
3.3	DATA WAREHOUSE	34
<b>3.3.1</b>	<b>Data Mart</b>	<b>36</b>
3.4	ETL	38
<b>3.4.1</b>	<b>Power Query Editor</b>	<b>39</b>
3.5	BUSINESS INTELLIGENCE	41
<b>3.5.1</b>	<b>Ferramenta <i>Microsoft Power BI</i></b>	<b>42</b>
3.6	INDICADORES DE DESEMPENHO	43
3.7	CONTROLADORIA	43
<b>4</b>	<b>ESPECIFICAÇÕES DO PROJETO</b>	<b>45</b>
4.1	MAPEAMENTO DO PROCESSO/ FLUXO DOS DADOS	45
4.2	REQUISITOS FUNCIONAIS	48
4.3	REQUISITOS NÃO FUNCIONAIS	48
4.4	PROPOSTA DE SOLUÇÃO	49
<b>5</b>	<b>DESENVOLVIMENTO</b>	<b>53</b>
5.1	ETAPAS DESENVOLVIMENTO	53
<b>5.1.1</b>	<b><i>Business Understanding</i></b>	<b>53</b>
<b>5.1.2</b>	<b><i>Data understanding</i></b>	<b>54</b>
<b>5.1.3</b>	<b><i>Data preparation</i></b>	<b>58</b>
<b>5.1.4</b>	<b><i>Modeling</i></b>	<b>69</b>
<b>5.1.5</b>	<b><i>Avaliação e Implantação</i></b>	<b>75</b>
<b>6</b>	<b>ANÁLISE DE RESULTADOS</b>	<b>76</b>
6.1	RESULTADO ARQUITETURA DE DADOS	76
6.2	RESULTADOS RELATÓRIOS <i>POWER BI</i>	78

<b>7</b>	<b>CONCLUSÃO . . . . .</b>	<b>80</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>82</b>
	<b>APÊNDICE A – CÓDIGOS M DAS ETAPAS DE CRIAÇÃO DAS TA- BELAS DO CONJUNTOS DE DADOS . . . . .</b>	<b>88</b>

## 1 INTRODUÇÃO

No mundo corporativo, as empresas seguem tendências para se manterem sempre atualizadas em um ambiente altamente competitivo. Nas últimas décadas existiu um investimento significativo em inovações (por meio de tecnologias) e grande parte deste investimento está vinculado à área de dados visando reduzir custos e melhorar a qualidade e o alcance dos sistemas de comunicação criando um paradigma de organização empresarial e de desenvolvimento econômico baseado na economia do conhecimento (SIQUEIRA, 2007).

Em (TEIXEIRA, 2017), cita-se textualmente:

*A possibilidade de estruturação de dados e o acesso a informações de qualidade em menor tempo vão liberar a inteligência financeira para o relacionamento com todas as áreas da organização e mesmo os públicos de interesse externos.*

Portanto, o fluxo de informações

deve ser organizado em uma arquitetura que pode ter diferentes etapas e especificações dependendo das ferramentas escolhidas para as funções desejadas visando a utilização na construção de conhecimento.

Este conhecimento é construído com o propósito de ser compartilhado e sua qualidade depende diretamente da modelagem da arquitetura que define o fluxo dos dados e seus requisitos.

Alguns exemplos de conhecimento baseado em dados são: objetivos estratégicos como ferramentas de aprendizado de máquina, geração de potenciais clientes, visualização de dados através de painéis interativos com tabelas e gráficos entre outros.

Para empresas do setor energético, como a Engie Brasil Participações, uma arquitetura de dados adequada é capaz de traduzir dados em informações para fomentar conhecimento como geração, compra, venda e preço. Estes conceitos são fundamentais para construir uma visão gerencial estratégica do negócio, acompanhar a situação da empresa, apoiar decisões, adquirir e compartilhar conhecimento.

O processo de *Business Intelligence*(BI), incluindo a arquitetura de dados, atual da empresa apresenta deficiências de escalabilidade, consistência, usabilidade, confiabilidade e automatização. Este processo de BI pode se nivelado como iniciante. A arquitetura atual é composta por arquivos coletados de diferentes tipos fontes e domínios. Estes são coletado de diferentes maneiras, diversas vezes necessitando da intervenção humana. Portanto, existem muitas inadequabilidades como reprocessamento de dados, redundância, vulnerabilidade, processos de *Extract, Transform and Load* (ETL), não documentados e descentralizados e arquivos locais. Os problemas decorrentes destas funções que divergem das necessidades do projeto, além de, em um

futuro próximo, apresentar problemas de escalabilidade, prejudicam a usabilidade da ferramenta e aumentar a vulnerabilidade do sistema a erros. Esses problemas colocam em risco o objetivo principal da área, uma vez que a tomada de decisões estratégicas baseada em dados incorretos é pior que a tomada de decisão sem levá-los em conta. Atualmente o sistema para transformar os dados usados no relatório final é mistificado, sendo difícil mapear fluxos para realização de auditorias.

## 1.1 OBJETIVOS

Considerando o cenário descrito na seção anterior, esta seção descreve o objetivo geral e os objetivos específicos deste trabalho.

### 1.1.1 Objetivo Geral

O objetivo geral do projeto é renovar a arquitetura de dados financeiros e de energia da área financeira de controle de risco da empresa Engie Brasil Participações. Desse modo, o fluxo de dados é automatizado, auditável e de fácil compreensão, eliminando ou diminuindo drasticamente a necessidade de interação humana no processo de geração dos dados. Diminuir a quantidade de interações humanas, torna este processo mais confiável. Esses objetivos visam, ao final, criar uma versão melhorada da visualização interativa da análise de dados estratégicos com visão de inteligência de negócio e da arquitetura de dados, apoiando a tomada de decisões da empresa através de monitoramento de cenários e entendimento compartilhado.

### 1.1.2 Objetivos Específicos

A partir do objetivo principal de construir uma arquitetura de dados estratégicos que seja confiável para a demonstrar a visão de inteligência de negócio, pode-se elencar os seguintes objetivos específicos:

- Replicação de um *Data Warehouse* contendo conjunto de dados de outras áreas da empresa utilizados no projeto;
- Integrar as diferentes fontes de dados para um mesmo repositório online;
- Criar o fluxo de transformações dos dados para disponibilizá-los para os relatórios;
- Modelar relacionamentos entre as diferentes tabelas na ferramenta final de visualização dos dados habilitando interação de dados entre diferentes tabelas;
- Automatizar o fluxo de dados do conjunto de a ser carregado na ferramenta final de inteligência de negócio;

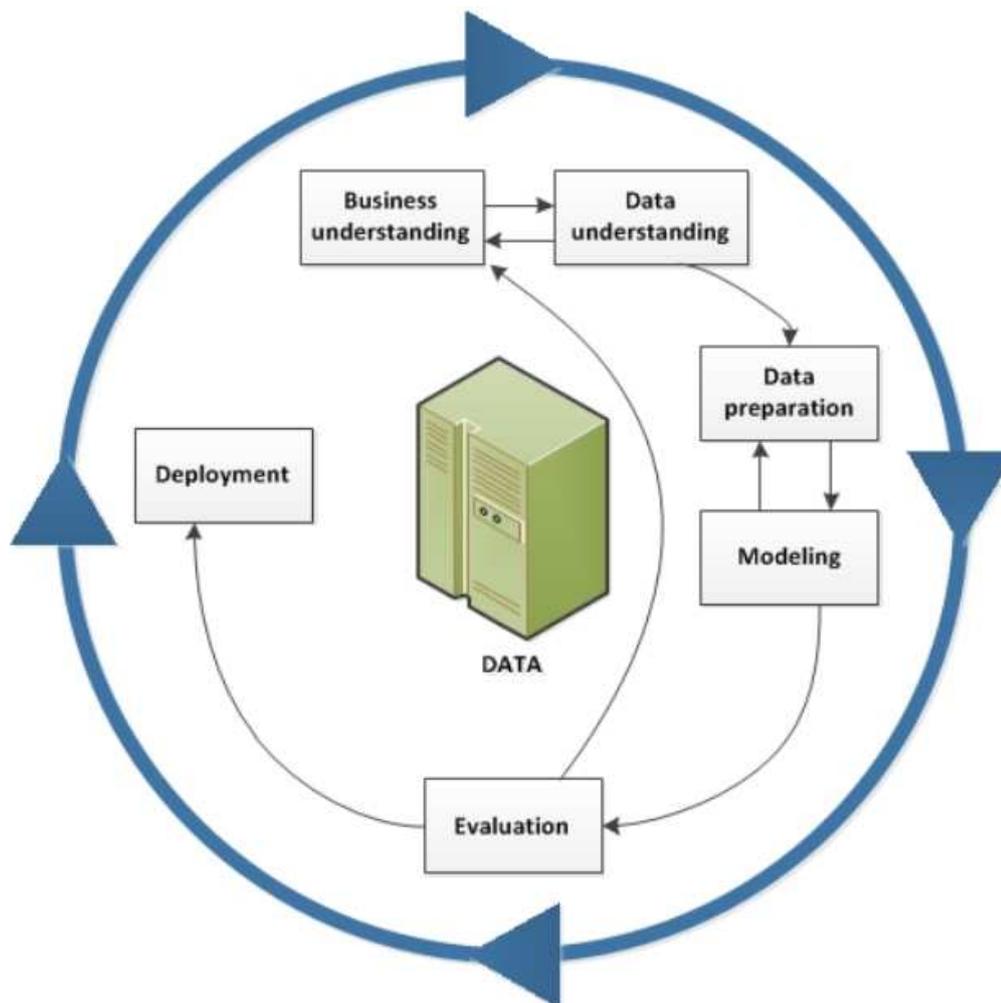
- Criar funções interativas e automatizações que auxiliem na construção dos relatórios;
- Construir relatórios paginados com performance e interatividade adequadas para área e os consumidores final do processo;
- Avaliar resultado final em comparação com arquitetura antiga.

## 1.2 METODOLOGIA

Neste projeto, a metodologia utilizada é a CRISP-DM (Cross Industry Standard Process for Data Mining), traduzida livremente para o português como Processo Padrão Interindústrias para Mineração de Dados, que é uma abordagem amplamente reconhecida pelo mercado para orientar no processo de mineração de dados (PIA-TETSKY, 2014).

Esta metodologia fornece uma visão geral do ciclo de vida da mineração de dados e é dividida em seis fases relacionadas. Estas fases podem ser revisitadas se necessário, seu uso beneficia o desenvolvimento do projeto tornando a implementação deste mais rápida, simples e fácil de gerenciar. Não sendo exclusivo uso de uma ferramenta ou de um tipo empresa (CHAPMAN *et al.*, 2000).

Figura 1 – Visão geral da ajuda do CRISP-DM.



Fonte: (IBM, 2024b).

Os principais relacionamentos das fases são indicados pelas setas na Figura 1, não sendo o sentido das setas uma regra mandatória. O modelo é customizável de acordo com o objetivo específico de cada ciclo. Este é estruturado e interativo, de modo que é possível a realimentação das fases diversas vezes até a conclusão do ciclo, alcançando o objetivo (IBM, 2024b).

Assim sendo, a metodologia pode ser utilizada para um projeto de arquitetura de dados para *Business Intelligence*, unindo objetivos estratégicos e poder analítico, gerando valor para a organização. Utilizando os conceitos das fases detalhadamente descritas por (CHAPMAN *et al.*, 2000) em *Step-by-step data mining guide* e os adaptando aos objetivos deste projeto, as fases e seus detalhamentos são:

1. Entendimento do negócio (*business understanding*): A primeira fase é de crucial para o projeto, portanto, investir tempo para definir um objetivo claro que considere as necessidades e o cenário atual da empresa é uma decisão inteligente

visando desenvolver uma solução que realmente gere valor para a organização. Nesta também deve-se formular hipóteses de como o problema será abordado levando em consideração os recursos disponíveis.

2. **Compreensão dos dados (*data understanding*):** Assim como a anterior, nesta fase de entendimento são formuladas hipóteses. O foco desta é coletar e entender os dados disponíveis, identificando não apenas as diferentes fontes e domínios, como também a qualidade, a construção e as características como tipo e granularidade. A qualidade destes levantamentos está diretamente ligada ao entendimento do negócio, juntas estas fases servem de base na formulação hipóteses para resolução do problema.
3. **Preparação dos dados (*data preparation*):** Nesta etapa o objetivo é, complementado com as duas fases anteriores, garantir que os dados sigam um padrão de formato adequado para análise. Esta fase engloba todas as atividades necessárias para consolidação da base de dados utilizada no projeto. Compõem esta fase atividades de transformação e processamento de dados utilizadas no projeto, como limpeza, construção, formatação, seleção, relação e filtragem entre outras. Assim como outras fases ela pode ser executada e modificada repetidas vezes, otimizando dados para obter cada vez um conjunto de dados mais robusto para objetivo principal.
4. **Modelagem (*modeling*):** Nesta fase, são criados modelos adequados para a hipótese levantada, considerando os dados e recursos disponíveis. Neste projeto, essa fase envolve como esses dados serão analisados para gerar valor em conhecimento estratégico. Isso implica em indicadores de desempenho, construção de possíveis cenários, métricas e automatização de rotinas serão utilizados.
5. **Avaliação (*evaluation*):** Essa fase possui como objetivo confirmar que os modelos desenvolvidos na fase anterior sejam válidos, relevantes e atingem os objetivos desejados do negócio. Nela ocorre a revisão do processo, avaliação de resultados. Por fim, deve-se concluir se o modelo está ou não de acordo com os requisitos para decidir as próximas ações.
6. **Implantação (*deployment*):** Esta é a fase final do processo, após validados os modelos pela fase de avaliação é realizada a construção e publicação de relatórios para monitoramento e manutenção, essas etapas finais garante o sucesso do projeto a longo prazo.

### 1.3 ESTRUTURA DO DOCUMENTO

Além deste capítulo, que apresenta uma introdução à motivação do projeto e à sua relevância, este documento é dividido em seis capítulos que são descritos a seguir.

O capítulo 2 apresenta o contexto no qual a empresa em que o projeto foi desenvolvido está envolvida, detalhando sua estrutura organizacional e setor, bem como o escopo da área financeira de controle de risco. Ainda apresenta a importância da arquitetura de dados, processos de inteligência de mercado no âmbito empresarial em questão.

No capítulo 3 é demonstrado o embasamento teórico utilizado para o desenvolvimento do projeto, e no capítulo 4 se detalha o processo AS-IS e TO-BE, a estrutura geral do projeto, requisitos funcionais e não funcionais, e a estrutura aplicada neste trabalho com auxílio de diagramas.

O capítulo 5 mostra o desenvolvimento da arquitetura e resultado final na ferramenta de BI. Nele estarão presentes os detalhes da integração, do fluxo de dados, das transformações e da construção da solução proposta, e os resultados obtidos serão expostos de uma forma aprofundada no capítulo 6.

Por fim, no capítulo 7, apresenta-se a síntese final do projeto como um todo.

## 2 EMPRESA ENGIE BRASIL PARTICIPAÇÕES

Figura 2 – Logo da empresa Engie Brasil.



Fonte: Institucional da empresa

A ENGIE Brasil Participações é *holding* do grupo franco-belga ENGIE no Brasil, essa empresa atua como controladora das entidades participantes do grupo no Brasil. Em território nacional ela é líder no setor de energia renovável privada, atuando em todo o Brasil, a empresa tem a totalidade de capacidade instalada proveniente de fontes limpas, renováveis e com baixa emissões de gases do efeito estufa no território como pode ser visto no mapa exibido na figura 4. A empresa atua no setor de energia em uma ampla gama de atividades como, geração, comercialização, mobilidade sustentável, soluções para cidades, transmissão de energia elétrica, transporte de gás e soluções energéticas. Fundada em 2008, a empresa possui escritórios no Brasil em Florianópolis e no Rio de Janeiro. Ela conta com uma capacidade instalada de cerca de 10 GW em 82 usinas representando aproximadamente 6 por cento da capacidade nacional. São mais de 3500 colaboradores no Brasil. Essas informações estão disponíveis através do painel exibido na figura 3, retirado do site da empresa.

Figura 3 – Painel de dados da ENGIE Brasil.

### ENGIE Brasil em números



Fonte: Institucional da empresa

Figura 4 – ENGIE no Brasil.



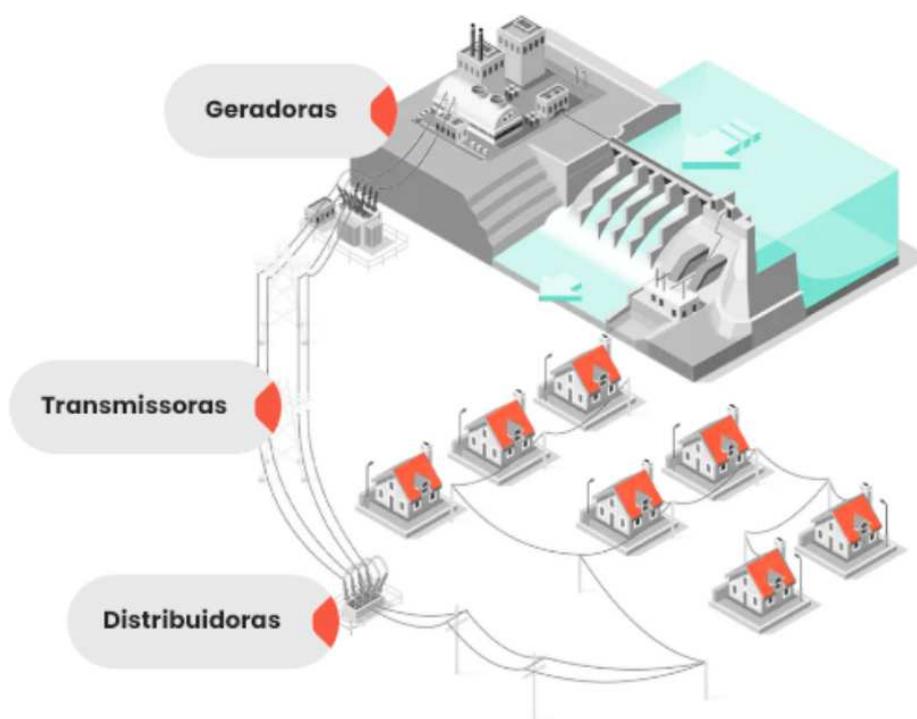
Fonte: Institucional da empresa

O grupo ENGIE possui forte presença global, atuando em diversos países e regiões ao redor do mundo. Ela tem uma ênfase na inovação e na busca de soluções energéticas mais sustentáveis, alinhando-se com a crescente conscientização sobre as questões ambientais e a busca por fontes de energia mais limpas e renováveis como os investimentos na construção de novas usinas eólicas e fotovoltaicas no nordeste do país. A ENGIE desempenha um papel importante no mercado global de energia e serviços, contribuindo para a transição energética para um futuro mais sustentável.

### 2.0.1 Setor de energia

o setor energético é vital para qualquer país, operacionalmente e financeiramente impactando diretamente na economia e desempenhando um papel fundamental na vida cotidiana das pessoas e no funcionamento de indústrias. (CASTRO; OSÓRIO; REGO, 2021) De maneira geral, a cadeia produtiva deste setor começa na geradoras, que são usinas que produzem energia para ser transportadas pelas linhas transmissoras que por fim são distribuídas pelas redes elétricas até o consumidor final sendo ele comum ou outras empresas (ABRADEE, 2024). Uma visão geral desta cadeia pode ser vista na figura 5

Figura 5 – Visão Geral da Cadeia setor energético.

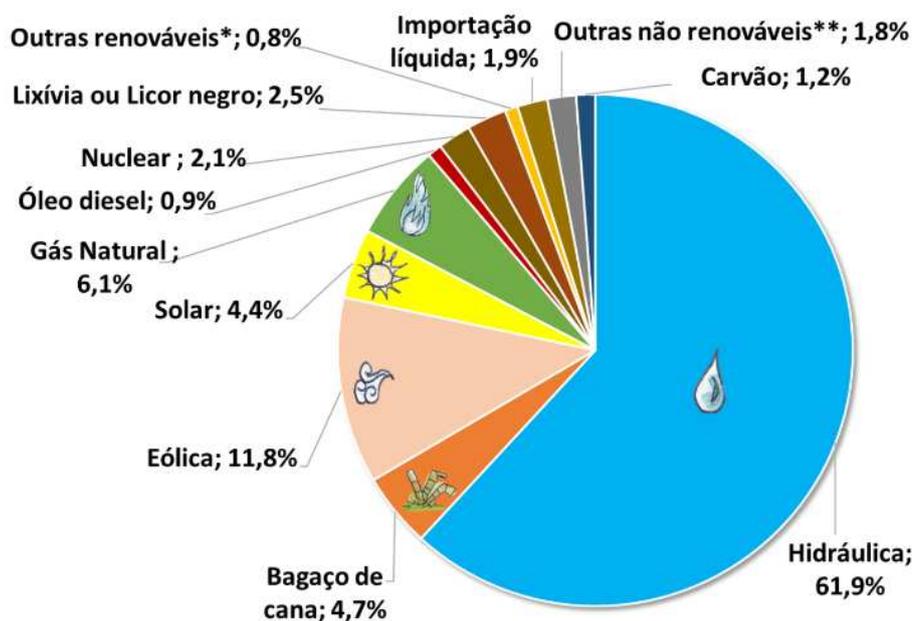


Fonte: (ABRADEE, 2024)

Analisando o mercado de energia do ponto de vista financeiro, é necessário reconhecer a complexidade das operações que envolvem entradas e saídas substanciais de capital, requisitando assim, um apurado gerenciamento de riscos para estabilidade e otimização de lucros. Para uma atuação satisfatória neste setor é vital gerenciar uma grande quantidade de ativos e passivos que podem ser traduzidos em compra, venda e geração de energia além de investimentos em infraestrutura como linhas de transmissão, operacionais, órgãos reguladores. O mercado de energia brasileiro é amplamente regulado pelo governo através do órgão governamental Agência Nacional de Energia Elétrica (ANEEL), de acordo com informações do site oficial da ANEEL (ENERGIA ELÉTRICA (ANEEL), 2024), vinculada ao ministério de Minas e Energia do país que também fiscaliza outros órgãos que fazem parte deste função como o Operador Nacional do Sistema Elétrico (ONS) órgão responsável pela coordenação e controle da operação da operação das instalações de geração e transmissão de energia elétrica no Sistema Interligado Nacional (SIN) e por planejar as operações dos sistemas isolados do Brasil. De acordo com informações do site oficial da (SISTEMA ELÉTRICO (ONS), 2024), a capacidade instalada de geração do sistema é predominantemente de hidrelétricas distribuídas em dezesseis bacias hidrográficas distribuídas no território nacional. A tendência nos últimos anos tem sido crescimento na instalação de usinas eólicas

agrupadas em sua maioria nas regiões Nordeste e Sul (OPERADOR NACIONAL DO SISTEMA ELÉTRICO (ONS), 2024).

Figura 6 – Matriz Elétrica Brasileira em 2022.



Fonte: (PESQUISA ENERGÉTICA (EPE), 2024)

Além das tecnologias já mencionadas também existe poucas usinas térmicas localizadas estrategicamente nos principais centros de carga. Estas usinas ajudam na segurança pois ajudam na gestão do estoque de água, aqui entendida como potencial de geração, nos reservatórios das usinas hidrelétricas assegurando a disponibilidade de atendimento. A matriz elétrica brasileira em 2022 é apresentada na figura 6 Essa centralização de gerenciamento junto com variedade de fontes favorece o sucesso de suprir o mercado consumidos com eficiência. Por outro lado, por ser regulado também reúne e divide informações do setor através de ferramentas e sites oficiais. A parte comercial que integra geradores, distribuidores, comercializadores e consumidos garantindo que a energia elétrica é a Câmara de Comercialização de Energia Elétrica (CCEE).

Como (VILELA, 2022) afirma em sua tese, o modelo vigente de comercialização de energia elétrica do Brasil possui dois ambientes, o Ambiente de Contratação Livre (ACL) e o Ambiente de Contratação Regulado (ACR). O ACR é composto por fontes geradoras e preço de contratação definidos pela ANEEL, é realizado através de leilões com o objetivo de atender necessidade de seus consumidores cativos, estes compram o preço de energia definido pela energia mais custosa compondo a matriz, chamado de Preço de Liquidação das Diferenças (PLD), apresentado na figura 7. O ambiente

de mercado livre, ACL, conta com consumidores livres e especiais, os quais diferente do regulado negociam de quem e por quanto compram sua energia.

Figura 7 – Painel de Preço de Liquidação das Diferenças no ambiente da CCEE.



Fonte: (CÂMARA DE COMERCIALIZAÇÃO DE ENERGIA ELÉTRICA, s.d.)

Uma das formas de comercialização da energia no ACL é através da plataforma Balcão Eletrônico de Comercialização de Energia (BBCE), essa plataforma digital disponibiliza um ambiente *online* de leilão dos ativos de energia elétrica. Ela permite que empresas cadastradas realizem, em tempo real, negócios de compra ou venda de energia através de ofertas. Essa plataforma também disponibiliza os dados das operações realizadas aos usuários, o que é utilizado para criar uma visão de mercado energético brasileiro (BALCÃO BRASILEIRO DE COMERCIALIZAÇÃO DE ENERGIA, s.d.).

## 2.0.2 Área financeira de controle de riscos

O grupo Engie Brasil é composto por algumas entidades como Soluções, Participações, Energia, Jirau. Este projeto foi desenvolvido na *holding* do grupo ENGIE no Brasil chamada de Engie Brasil Participações (EBP). Sucintamente essa *holding* é a empresa matriz das subsidiárias no Brasil com relação direta com a fundadora francesa, sendo responsável por esse vínculo, por essa razão tem um caráter de controladoria participando da política e gestão interna de cada uma das empresas subsidiárias como pode ser visto no diagrama organizacional na figura 8.

O conjunto de dados utilizados pela área refletem o contexto de preço e volume de energia. Estes estão relacionados entre si e possuem diversas subcategorias usadas para análises estratégicas servindo de embasamento para decisões de ditam o

Figura 8 – Diagrama organizacional ENGIE Brasil.



Fonte: Aluno.

futuro da empresa e seus funcionários. Como (ZDANOWICZ, 1989) apresenta em seu livro, "o Fluxo de Caixa: uma decisão de planejamento e controle financeiro", a importância da análise de dados financeiros de modo que a administração financeira gere resultados positivos nas operações de caixa garantindo condições de lucros que compensem a exposição aos riscos. Como mencionado na seção anterior o mercado de energia é muito complexo, e é comum deparar-se com estimativas futuras de momentos passados, como o cenário esperado de geração no mês anterior ou ano anterior, tornando necessário colocar-se em um momento no passado enquanto olha um cenário futuro<sup>1</sup>. Um exemplo simples é a venda de energia, nela a empresa precisa estimar volume disponível e principalmente qual será o preço da energia meses ou anos no futuro. Essa visão muda conforme o tempo ou versão dos dados que representam cenários empresariais. Um esquema é apresentado na Fig. 9.

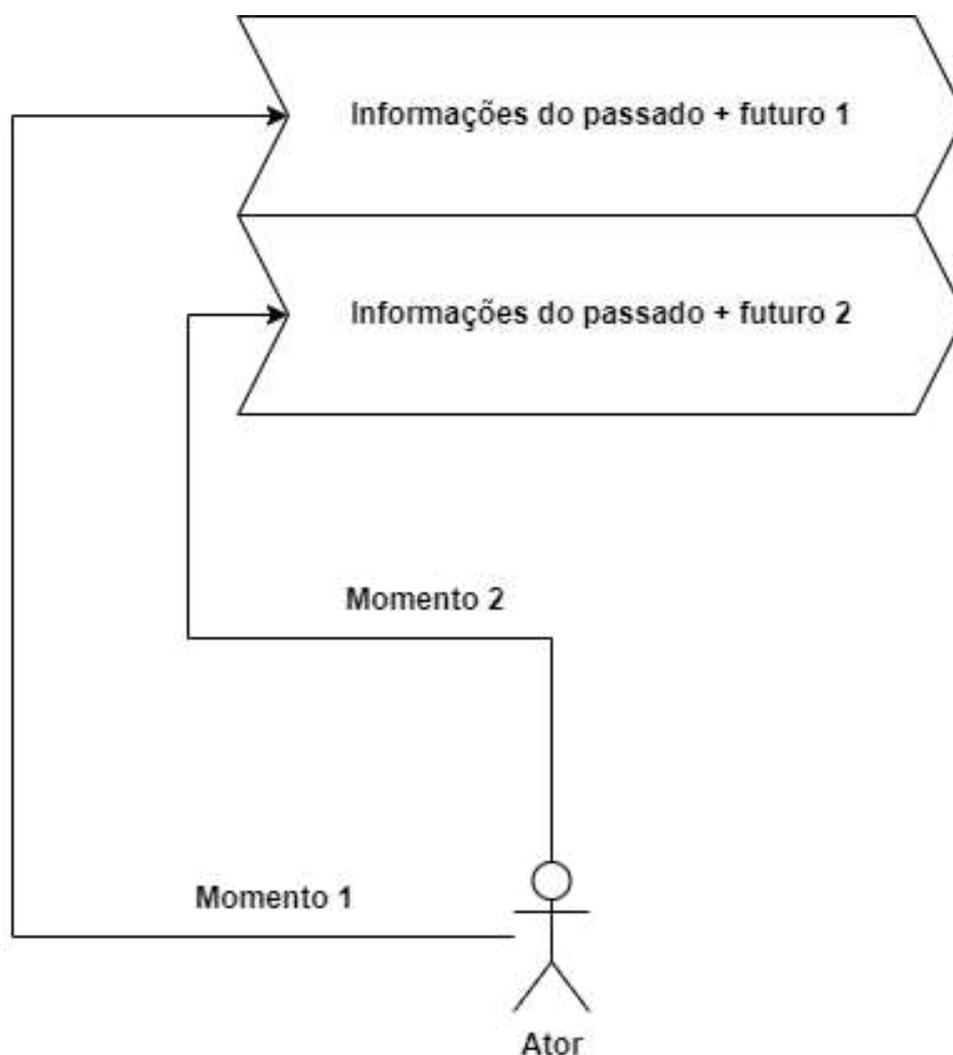
O departamento financeiro pode ser responsável pelo gerenciamento dos riscos de mercado, câmbio, crédito e operacional da própria empresa. Este gerenciamento deve ser realizado de forma a antecipar riscos, para que as medidas adequadas possam ser tomadas em tempo oportuno evitando problemas ou reduzindo os seus impactos.

A equipe de controle de risco da área financeira desempenha um papel de grande impacto na empresa, sendo ela responsável pelo gerenciamento de riscos do grupo Engie no Brasil. Entende-se por gerenciamento de riscos como o processo de identificar, monitorar, analisar e controlar os riscos que envolvem o negócio.

As empresas que atuam nesse setor precisam gerenciar com excelência a parte financeira para manter a continuidade das operações e o fornecimento confiável de energia à sociedade. Essa abordagem supervisória e estratégica é feita pela área de

<sup>1</sup> Nas análises da área, utilizam-se dados estimados do futuro que são calculados por outras áreas e não se consideram modelos matemáticos complexos.

Figura 9 – Visão dos dados de energia.



Fonte: Aluno.

controle de risco e suas atividades são apresentadas a seguir:

- Criação e acompanhamentos de KPIs;
- Elaboração de relatórios personalizados através do Power BI para análises embutidas;
- Participação da estruturação, aquisição, integração e transformação de dados;
- Criação de análises a partir de dados de clima, financeiros e de mercado;
- Criação e atualização no desenvolvimento de indicadores de desempenho por tecnologia.

- Controle do desempenho das tecnologias e impacto no resultado financeiro da empresa.
- Elaboração de apresentações mensais da área de controle de riscos de mercado e de crédito para diretores e cargos de chefias.
- Avaliação de Risco de Comercialização de Planta Geradora do Grupo ENGIE para melhor relação Risco vs Retorno.

Estas atividades compõem o dia a dia dos colaboradores desta área mantendo o relatório atualizado e disponível para eventuais consultas de líderes, diretores e de colaboradores outras áreas, elaborar e apresentar a reunião mensal com dados mês a mês adicionando informações e análises importantes, além de prestar apoio e auditoria de outras áreas da empresa.

### 3 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados os aspectos conceituais teóricos utilizados no desenvolvimento da solução que envolve a modelagem de uma arquitetura de dados para processo de *Business Intelligence* (BI). Também serão elucidadas as ferramentas e linguagens de programação, que são a base dos capítulos seguintes.

#### 3.1 BANCO DE DADOS

Nesta seção serão abordados os conceitos, modelos e técnicas as quais fazem parte da fundamentação teórica de banco de dados utilizado no projeto. Assim como a justificativa do uso desses fundamentos para o desenvolvimento da solução do problema abordado.

*"Os primeiros bancos de dados eram fitas magnéticas com registros de dados armazenados sequencialmente. Os bancos de dados continuaram a evoluir com os avanços da tecnologia. Eles agora se tornaram sistemas complexos e de alta performance com seu próprio campo de estudo dedicado"(AMAZON WEB SERVICES, 2024).*

Figura 10 – Modelo de Banco de dados.



Fonte: (AMAZON WEB SERVICES, 2024).

Para O'Brien (O'BRIEN, 2010), os conceitos de dados e informações é diversas vezes confundido. Dados são entendidos como matéria-prima para o processo que constrói informação. De acordo com a definição da empresa *Oracle*, um banco de dados é uma coleção organizada de dados estruturados, usualmente armazenados em um sistema de computador e controlado por um sistema de gerenciamento de bando de dados (DBMS), ele é integrado a diversas ferramentas e aplicativos criando um sistema de banco de dados (ORACLE, 2024). Os sistemas de bancos de dados são utilizados para armazenar e organizar dados de maneira gerenciável e acessível. Todas as ações

realizadas são registradas gerando dados, esses são transformados em informações. Alguns exemplos mais comuns de bancos de dados são de transações financeiras, catálogos de comércio eletrônico e redes sociais (MICROSOFT AZURE, 2024). Um banco de dados adequado é muito importante para todas as organizações, levando em consideração necessidades como escalabilidade, eficiência, integridade, segurança e análises. Estes bancos de dados podem servir de fonte para operações internas como informações pessoais dos colaboradores da empresa, dados de operação de geração de energia ou externas como interação com clientes (AMAZON WEB SERVICES, 2024).

### 3.1.1 Banco de dados Relacional

Segundo site oficial de uma das empresas dominantes no mercado de tecnologia relacionado a dados, um modelo de banco de dados mostra a estrutura lógica de um banco de dados. Nele são define os relacionamentos e as regras que determinam como os dados podem ser armazenados, organizados e manipulados. Cada aplicação de banco de dados é construída em um modelo de dados específico de acordo com os requisitos para atender o objetivo final (AMAZON WEB SERVICES, 2024).

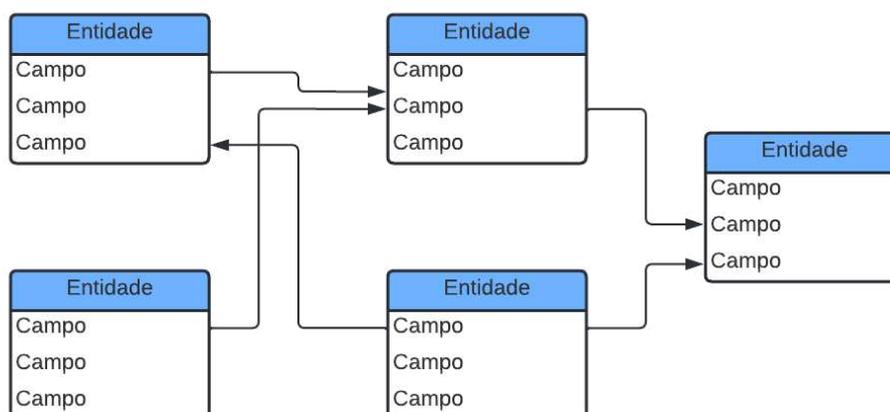
O conceito de banco de dados relacional é antigo, em 1970, o pesquisador (CODD, 1970) publicou um artigo que teoriza o modelo de bando de dados relacional. Neste artigo ele propôs organizar dados com base em relacionamento significativos como tuplas ou pares atributo-valor. Estes foram chamados de relações permitindo fusão de dados entre tabelas.

Este modelo de banco de dados é amplamente utilizado pelo fato de ser amigável ao uso de pessoas sem formação na área de tecnologia. Os tipos de relacionamentos entre os dados são listados abaixo conforme documentação da IBM, sendo os mais utilizados, o um para um e um para muitos (IBM, s.d.).

Citado em (RAMAKRISHNAN; GEHRKE, 2011):

*“As principais vantagens do modelo relacional em relação aos modelos de dados mais antigos são sua representação de dados simples e a facilidade com que mesmo consultas complexas podem ser expressas.”*

Figura 11 – Modelo genérico banco de dados relacional.



Fonte: (BARBOSA FILHO, 2022).

Um exemplo do modelo de dados relacional pode ser visualizado na figura 11 foi amplamente difundido devido a vantagem de ser flexível e suprir as necessidades de desafios como concepção e implementação do sistema de base de dados. Como o próprio nome já diz a estrutura fundamental do modelo é a relação, ela assume que os dados já estão devidamente normalizados. Nesse modelo as colunas de uma tabela como um atributo e cada linha como sendo um registro ou instância do esquema, sendo assim cada linha deve possuir uma informação por coluna. Estes atributos também possuem rótulo que define o tipo de dado como data, número, texto entre outros (BONFIOLI, 2006). É possível ver um exemplo de atributos na tabela 12

Figura 12 – Atributos e Registros de banco de dados relacionais.

ID	Atributo1	Atributo2	Atributo3	Atributo4
	Nome	Idade	Gênero	cpf
Registro1	Alex	27	Masc	111.111.111-11
Registro2	Funalo1	18	Masc	222.222.222-22
Registro3	Funalo2	19	Masc	222.222.222-23
Registro4	Funalo3	20	Masc	222.222.222-24
Registro5	Funalo4	21	Masc	222.222.222-25

Fonte: Aluno.

Com as tabelas devidamente tratadas, são construídos os relacionamentos entre os atributos destas com o cada tipo de acordo com suas características. Nor-

malmente estes são utilizados com objetivo de evitar redundância entre os dados melhorando a performance e de facilitar o entendimento das informações contidas no conjunto de dados. Algumas das funções da construção desses relacionamentos utilizadas no projeto são de: Mapeamento de atributos comuns a diferentes tabelas com indexação e para filtros nos visuais, tabelas como de datas possibilitando utilizar outras formatações e interação de tabelas diferentes para análises. Também é utilizado o conceito de chaves primárias que são únicas identificando entidades únicas para aquela informação. Os tipos de relacionamentos são apresentados na tabela 13

Figura 13 – Tipos de relacionamentos de banco de dados.

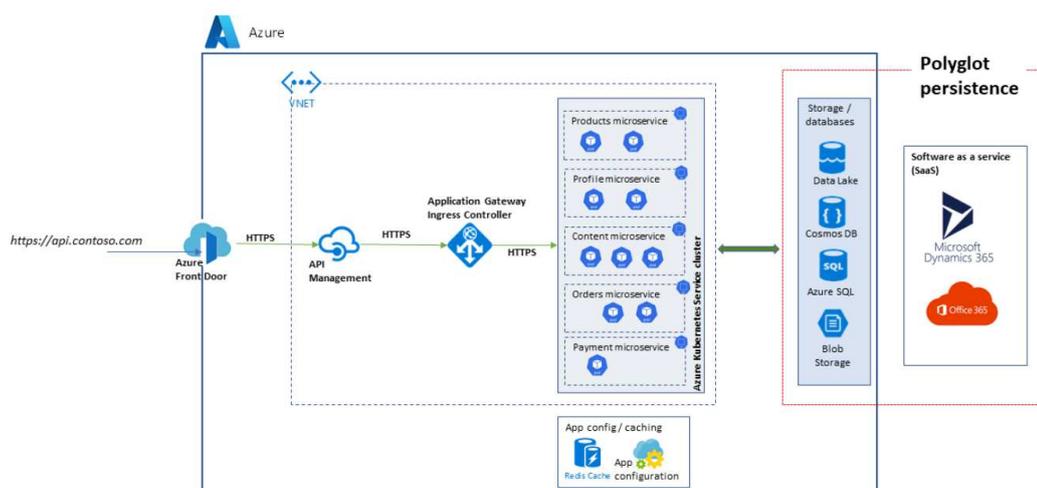
Tipo de relacionamento	Descrição
Uma para um	Ambas tabelas podem ter somente um registro de cada lado do relacionamento. Cada valor da chave primária se relaciona a nenhum ou a apenas um registro na tabela relacionada. A maioria dos relacionamentos de um para um são forçados por regras de negócios e não fluem naturalmente dos dados. Sem tal regra, geralmente você pode combinar as duas tabelas sem quebrar nenhuma regra de normalização.
Uma para muitos	A tabela de chave primária contém somente um registro relacionado a nenhum, a um ou a muitos registros da tabela relacionada.
Muitos para muitos	Cada registro em ambas as tabelas pode se relacionar a nenhum ou a qualquer número de registros na outra tabela. Esses relacionamentos requerem uma terceira tabela, chamada de tabela associada ou de associação, pois os sistemas relacionais não podem acomodar diretamente o relacionamento.

Fonte: (IBM, 2024a).

## 3.2 TECNOLOGIA NA NUVEM

De acordo com uma das grandes empresas consolidada no setor de tecnologia Amazon, a definição de computação em nuvem é a entrega de recursos de TI sob demanda por meio da internet com definição de preço e pagamento conforme o uso (AMAZON, s.d.). No projeto a infraestrutura é fornecida por terceiros pela Microsoft Azure, os usuários configuram remotamente e criam ou instalam suas soluções nela. A infraestrutura em nuvem de nível empresarial executa sistemas críticos com confiança e segurança, ela é composta por dois componentes principais: a infraestrutura física, que são edifícios físicos exclusivos localizados ao redor do mundo e abrigam um grupo de servidores de computador em rede e é neutra em carbono desde 2012, e os componentes de rede de conexão, nele os aplicativos são integrados como mostra o esquema genérico da nuvem na figura 14 (MICROSOFT AZURE, s.d.).

Figura 14 – Esquema Nuvem da Microsoft.



Fonte: (MICROSOFT, s.d.[a]).

O armazenamento em nuvem é um conjunto de serviços, ele permite armazenar e compartilhar dados ao migrá-los via rede a um sistema de armazenamento externo terceirizado de forma segura e eficiente (AMAZON, s.d.). Esse sistema usa *software* como serviço é composto por aplicativos executados no ambiente em nuvem. Eles ficam disponíveis através de uma interface online acessível por diversos usuários através de dispositivos com acesso. Assim, ele permite que novos recursos sejam adicionados aos sistemas de forma que todos os usuários fiquem cientes das mudanças, tornando a manutenção e evolução dos sistemas de conhecimento geral (BORGES *et al.*, 2011).

Os principais benefícios que justificam o uso desse conceito no projeto, de acordo com documentação da Microsoft (AMAZON, s.d.), são:

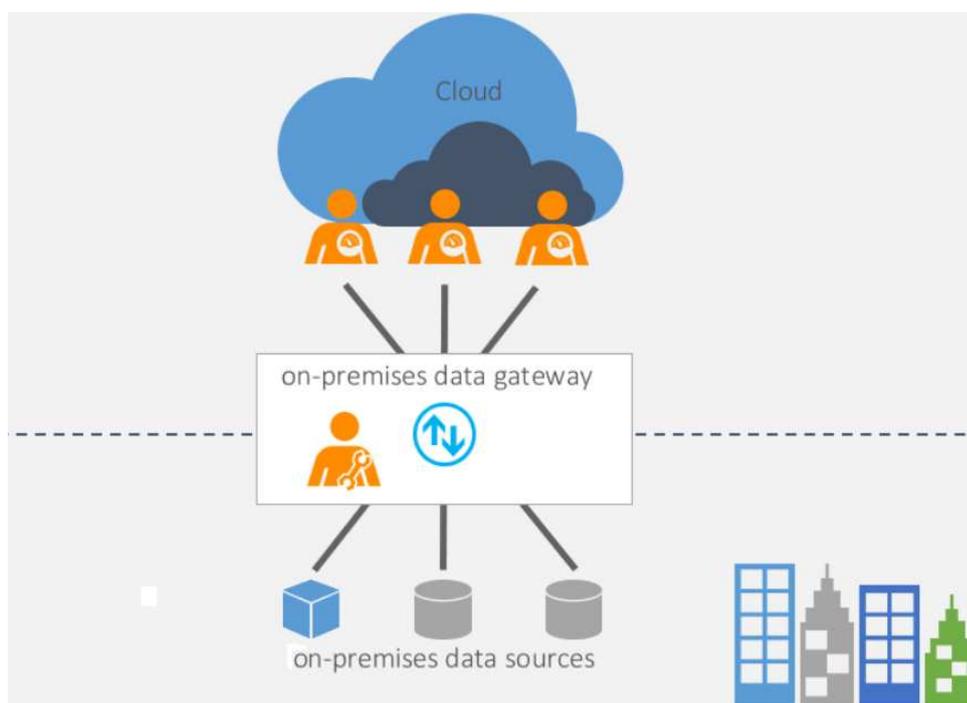
- **Custo:** Diminui dos custos de tecnologia substituindo a compra de hardware e software físicos pela contratação do conjunto de serviços, gerenciando a infraestrutura terceirizada por especialistas entre outros;
- **Velocidade:** Fornece flexibilidade e rapidez pelo fato de a maior parte dos serviços de computação em nuvem serem fornecidos sob demanda por autosserviços;
- **Escalabilidade:** Fornece os requisitos de TI necessários para os serviços, tornando-se assim adaptável de acordo com o dimensionamento;
- **Produtividade:** Dispensa a necessidade de diversas tarefas dos sistemas físicos como configuração e correção. Assim aumenta a produtividade liberando tempo de trabalho dos profissionais que estariam envolvidos;

- Desempenho: O fato de o serviço ser especializado implica na atualização constante com as melhores soluções. Devido a isso os serviços têm latência de rede reduzida para aplicativo e economia de escalonamento.

### 3.2.1 Gateway de dados

O gateway de dados local funciona como uma ponte. Ele é responsável pela transferência de dados rápida e segura, permitindo que usuários conectem-se a diversas fontes de dados através serviços em nuvem da Microsoft. Esses serviços incluem *Power BI*, *Azure Analysis Services* e Aplicativos Lógicos do Azure (MICROSOFT, s.d.[b]). O principal benefício que justifica a utilização de um gateway para as organizações é manter os bancos de dados e outras fontes de dados em suas redes locais em paralelo com o uso desses dados locais com segurança em serviços de nuvem (MICROSOFT, s.d.[b]). Em outras palavras no projeto é necessário para atualização automática dos dados e relatório. Um exemplo do esquema de ligação de um gateway é apresentado na figura 15.

Figura 15 – Esquema gateway.



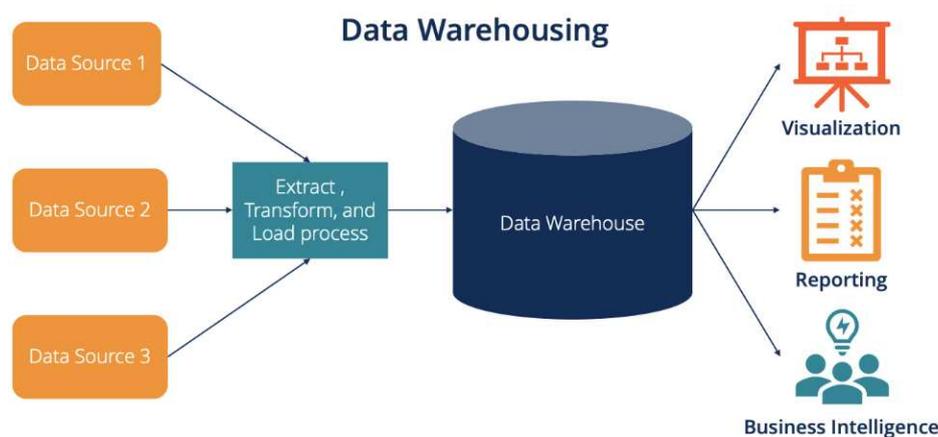
Fonte: (MICROSOFT, s.d.[b]).

## 3.3 DATA WAREHOUSE

O conceito de *Data Warehouse (DW)* segundo (TURBAN *et al.*, 2009), pode ser definido como um conjunto de técnicas e banco de dados integrados, projetado para

executar as funções do sistema que disponibiliza dados para formação de conhecimentos que servem para apoio a decisões estratégicas, como apresentado na figura 16.

Figura 16 – Data Warehouse genérico.



Fonte: (CORPORATE FINANCE INSTITUTE, s.d.).

Para (KIMBALL, 2002), *Data Warehouse* é o conglomerado de todos os *Data Marts* dentro da empresa utilizados para determinado setor. O *Data Mart* seria a própria unidade lógica do *Data Warehouse*.

Na tabela 1, (ANDREATTO, 1999) deixa claro as diferenças entre banco de dados operacionais e *Data Warehouse* com base em características relevantes.

No livro de sistemas de banco de dados (NAVATHE; ELMASRI, 2005) é defendido o uso do conceito de *Data Mining* e *Knowledge Discovery* (KDD) como complemento do uso de um *Data Warehouse*:

*A Descoberta de Conhecimento em Bancos de Dados (Knowledge Discovery in Databases), normalmente abreviada como KDD, engloba mais que a Data Mining. O processo é composto por seis fases: seleção de dados, limpeza, enriquecimento, transformação ou codificação, Data Mining e construção de relatórios e apresentação da informação descoberta.*

A mineração de dados tem como objetivo extrair conhecimento a partir de um grande conjunto de dados, descobrir padrões e relações desconhecidas para relacionar informações de modo que sejam úteis para análises estratégicas auxiliando na tomada de decisões gerenciais embasadas (PEREIRA, 2021).

O DW armazena dados analíticos, destinados às necessidades da gerência no processo de tomada de decisões. Isto pode envolver consultas complexas que necessitam acessar um grande número de registros, por isso é importante a existência

Tabela 1 – Tabela comparativa Banco de dados vs Data Warehouse.

Características	Bancos de dados Operacionais	Data Warehouse
Objetivo	Operações diárias do negócio	Analisar o negócio
Uso	Operacional	Informativo
Unidade de trabalho	Inclusão, alteração, exclusão.	Carga e consulta
Número de usuários	Milhares	Centenas
Tipo de usuário	Operadores	Comunidade gerencial
Interação do usuário	Somente pré-definida	Pré-definida e ad-hoc
Condições dos dados	Dados operacionais	Dados Analíticos
Volume	Megabytes – gigabytes	Gigabytes – terabytes
Histórico	60 a 90 dias	5 a 10 anos
Granularidade	Detalhados	Detalhados e resumidos
Redundância	Não ocorre	Ocorre
Estrutura	Estática	Variável
Manutenção desejada	Mínima	Constante
Acesso a registros	Dezenas	Milhares
Atualização	Contínua (tempo real)	Periódica (em batch)
Integridade	Transação	A cada atualização
Número de índices	Poucos/simples	Muitos/complexos
Intenção dos índices	Localizar um registro	Aperfeiçoar consultas

Fonte: (ANDREATTO, 1999).

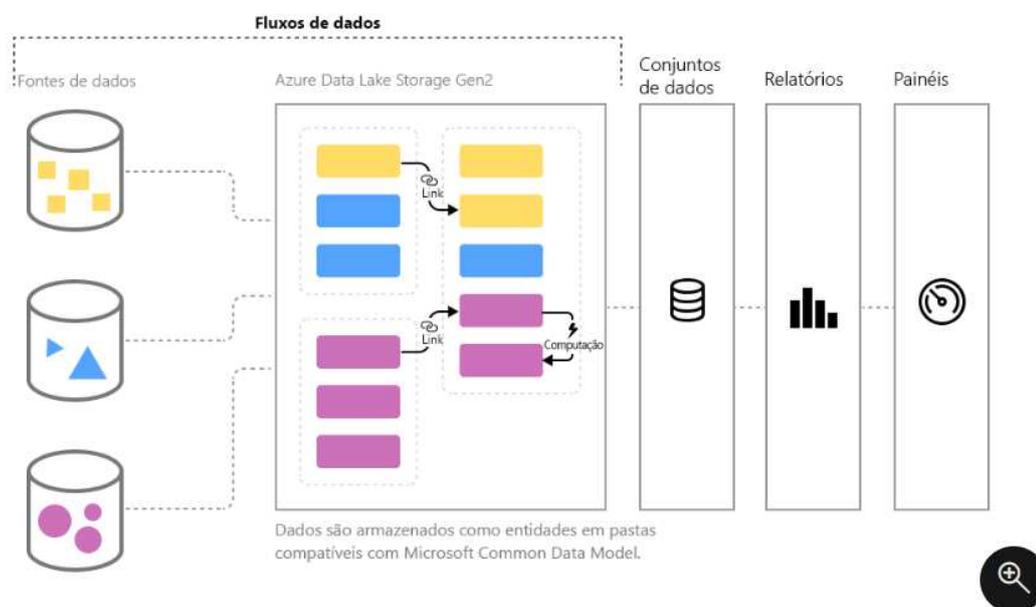
de muitos índices criados para acessar as informações da maneira mais rápida possível (PRIMAK, Fabio Vinícius, 2008).

### 3.3.1 Data Mart

O termo *Data Mart* (DM) refere-se a um depósito de dados relacionados a um setor específico da empresa, como financeiro por exemplo, desenvolvidos com objetivo de auxiliar no processo decisório do setor da empresa. É possível abstrair o conceito de *Data Mart* como sendo um *Data Warehouse* setorial agrupando informações enriquecidas com análises necessárias (JESUS, A. C. d., 2020). De acordo com (JESUS, A. C. d., 2020), o *Data Mart* é um subconjunto de dados pertencentes as áreas específicas e delimitadas de dados. Este funciona de forma a acelerar o processo de consulta de informação da área dispensando a necessidade da visão de dados desnecessários utilizados por outros setores da empresa e podendo ser gerenciados por profissionais não especializados. Estes dados são extraídos de uma *Data Warehouse* ou de outras fontes como pastas locais ou em nuvens, outros.

Os *Data Marts* são construídos para responder de forma embasada em informações possíveis questionamento de um grupo de outro setor ou tipo específico de usuário. Neste projeto, ele se destinará a fornecer um panorama geral, planejamento de cenários e controle do grupo ENGIE e entidades participantes focado nas informações do Brasil.

Figura 17 – Esquema estrutura Data Marts.



Fonte: (MICROSOFT, s.d.[d]).

No *Data Mart* os dados são extraídos, carregados e transformados gerando automaticamente um modelo semântico, ou um conjunto de dados utilizado de base para criar relatórios e *dashboards* do Power BI. Essa estrutura é apresentada na figura 17. Dentro dos benefícios, estão a facilidade com que usuário podem utilizar o sistema rompendo a dependência de um profissional de TI para qualquer atividade de análise de dados, dando a esses outros profissionais autossuficiência. Oferecem também a vantagem de processo ponta a ponta na preparação, exploração utilizando conceitos de SQL, mas em uma ferramenta que possibilita baixa necessidade uso de código (MICROSOFT, s.d.[d]).

Este conceito foi projetado para dar suporte a diferentes tipos de cenários, a tabela 18 retirada da documentação *Microsoft* demonstra o melhor uso de cada uma dela e suas funções:

Figura 18 – Tabela Funções Data Mart.

Item	Caso de uso recomendado	Complementando a função com os datamarts
Datamarts	Armazenamento de dados baseado no usuário e acesso SQL aos seus dados	Os datamarts podem ser usados como fontes para outros datamarts ou itens, por meio do ponto de extremidade SQL: <ul style="list-style-type: none"> <li>• Compartilhamento externo</li> <li>• Compartilhamento entre os limites departamentais ou organizacionais com segurança habilitada</li> </ul>
Fluxos de dados	ETL (preparação de dados reutilizáveis) para modelos semânticos ou marts	Os datamarts usam apenas um fluxo de dados interno para ETL. Os fluxos de dados podem acentuar isso, habilitando o seguinte: <ul style="list-style-type: none"> <li>• Carga de dados em datamarts com diferentes agendamentos de atualização</li> <li>• Separação das etapas de preparação de dados e ETL do armazenamento, para que possam ser reutilizados pelos modelos semânticos</li> </ul>
Modelos semânticos	Métricas e camada semântica para relatórios de BI	Os datamarts fornecem um modelo semântico gerado automaticamente para relatórios, habilitando: <ul style="list-style-type: none"> <li>• Combinar dados de várias fontes</li> <li>• Compartilhamento seletivo das tabelas do datamart para relatórios refinados</li> <li>• Modelos compostos: modelo semântico com os dados do datamart e de outras fontes de dados externas</li> <li>• Modelos de proxy: um modelo semântico que usa DirectQuery para o modelo gerado automaticamente, com apenas uma fonte de verdade</li> </ul>

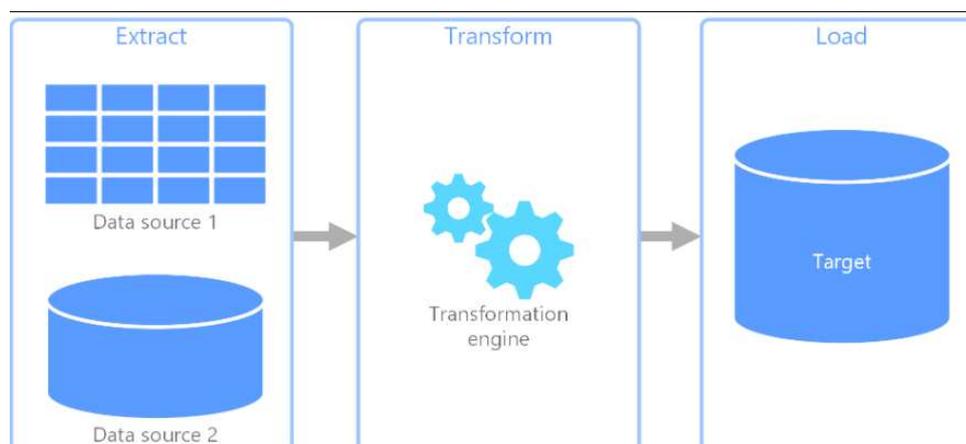
Fonte: (MICROSOFT, s.d.[d]).

### 3.4 ETL

Dentre os desafios corriqueiros, presentes nas organizações em geral, estão: coletar dados de diversas fontes em diferentes formatos e não estruturados que são inadequados para análises. Um desafio comum também é de armazenar esses dados em um repositório sem perder versões antigas. Diferentes ferramentas, serviços e processos foram desenvolvidos ao longo dos anos afim de solucionar esses desafios. Independente do processo de formação de conhecimento é evidente necessidade de gerenciar o fluxo de dados na arquitetura (MICROSOFT, 2024).

ETL (*Extract, Transform and Load*), traduzido do inglês como três verbos: extrair, transformar e carregar, é um processo executado por ferramentas de *software* que desempenha função resumida pelos verbos que compõem a sua nomenclatura. Assim este processo extrai dados de uma ou mais fontes de diferentes origens criando uma cópia dos dados extraídos (JESUS, A. C. de, 2020).

Figura 19 – Esquema Processo ETL.



Fonte: (MICROSOFT, 2024).

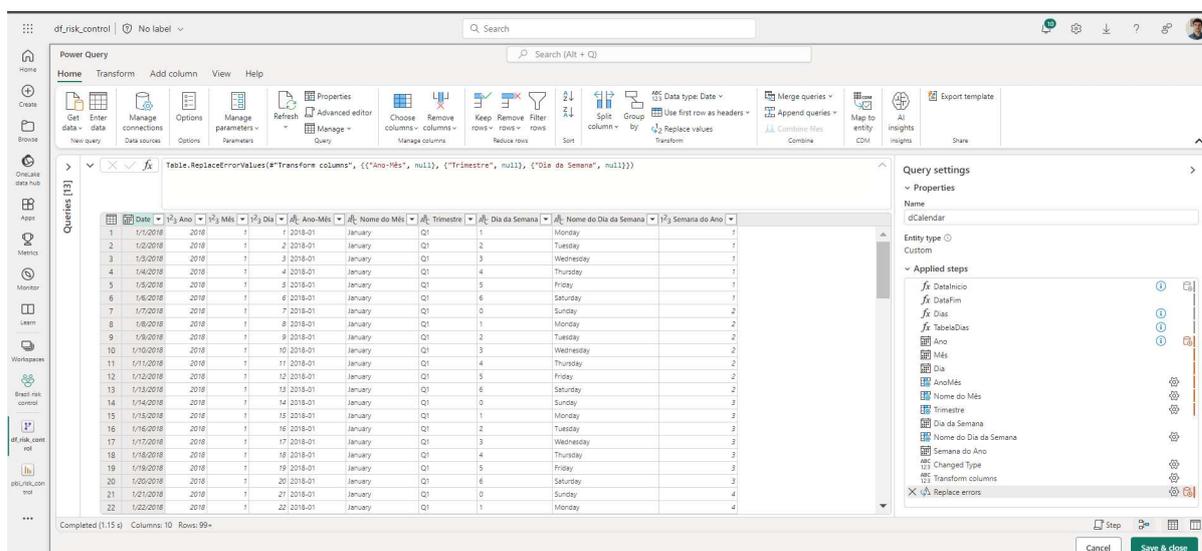
Após reunir dados de distintos tipos de fontes como planilhas Excel, arquivos de textos, *Data Warehouse* entre outros para o modelo semântico a próxima etapa é de transformação, nesta os dados extraídos, estruturados ou não, são processados afim de realizar limpeza, construção de análises, padronização, rotulagem dos dados garantindo integração e confiabilidade (KIMBALL, 2002). O processo de ETL é controlado utilizando fluxo de dados, sendo possível adicionar visualizar etapas intermediárias conforme eles são processados (MICROSOFT, 2024). Ao final esses dados processados são carregados na ferramenta final de *Business Intelligence*, como mostra a figura 19, para compartilhamento de conhecimento. Esse processo pode ser agendado conforme a necessidade do projeto.

### 3.4.1 Power Query Editor

Segundo definição da empresa proprietária da ferramenta *Power Query Editor*, a *Microsoft*: (MICROSOFT, s.d.[c])

*"O Power Query é um mecanismo de transformação e preparação de dados. O Power Query vem com uma interface gráfica para obter dados de fontes e um editor do Power Query para aplicar transformações. Como o mecanismo está disponível em muitos produtos e serviços, o destino no qual os dados serão armazenados depende de onde o Power Query foi usado. Usando o Power Query, você pode executar o processamento ETL (extrair, transformar e carregar) com os dados."*

Figura 20 – Interface da ferramenta Power Query.



Fonte: Aluno.

A linguagem de consulta estruturada *Sequence Query Language* (SQL) foi inventada por Don Chamberlin e Ray Boyce na IBM. Ela é a linguagem padrão para interagir com sistemas de gerenciamento de banco de dados relacionais com objetivo principais de consultar, inserir, atualizar, excluir, editar (IBM, 2024a).

Utilizando conceitos de SQL e linguagem de *script* foi desenvolvida para a ferramenta Power Query Editor a linguagem de fórmula M, também conhecida apenas por linguagem M. Ela foi criada para tornar mais amigável a programação aos usuários, o processo pode ser criado quase sem necessariamente escrever um código enquanto é escrita automaticamente nos bastidores de toda interação do usuário na ferramenta. Também pode-se editar diretamente o código aumentando as possibilidades do usuário (IBM, 2024a).

Exemplo:

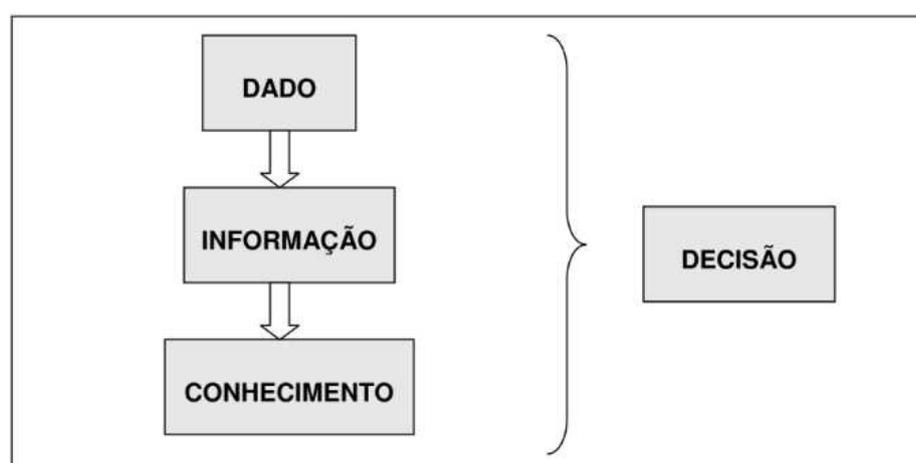
- Nome\_da\_etapa = Função(Etapa\_carregado, each ([Coluna] <> "filtro"))

Essa linguagem é construída em etapas, como mostra o exemplo acima da construção de uma etapa, a primeira é a de extração dos dados, podendo ser vinculada a diversos tipos de fontes de dados, inclusive próprios dados já carregados. O resultado parcial de cada uma destas etapas de ETL podem ser consultadas no meio do processo. Ela segue um padrão de fórmula que primeiro referência a etapa fonte para depois realizar ações determinadas. Por fim a etapa desejada é carregada para o *Power BI*. A interface do Power Query pode ser vista na figura 20.

### 3.5 BUSINESS INTELLIGENCE

O *Business Intelligence* é um sistema que agrupa informações de uma empresa, ele é composto por fases de coleta, análise, interação e validação destas informações agrupadas. Validadas, essas informações são analisadas com o objetivo de traduzi-las em conhecimento estratégico embasando decisões gerenciais (SERIAN, 2007).

Figura 21 – Dado x Informação x Conhecimento.



Fonte: (PRIMAK, Fábio Vinícius, 2008).

Atualmente existe a necessidade de um sistema de BI não importa o porte da organização. Essa construção de conhecimento auxilia diferentes cenários guiando gestores para uma decisão assertiva, reduzindo riscos, custos, otimizando processos, planejando e elaborando estratégias (PRIMAK, Fábio Vinícius, 2008). Essas aplicações de BI podem ser entregues usando *softwares* de código aberto e demandas via formulários online como um serviço de software (LORENZETTI, 2010).

A ênfase do *Business Intelligence* é na tomada de decisão eficaz, entretanto é importante levar em conta o embasamento utilizado para a tomada de decisão no âmbito da organização. (PRIMAK, Fábio Vinícius, 2008) afirma em seu livro, *Decisões com BI (Business Intelligence)*: (MICROSOFT, s.d.[c])

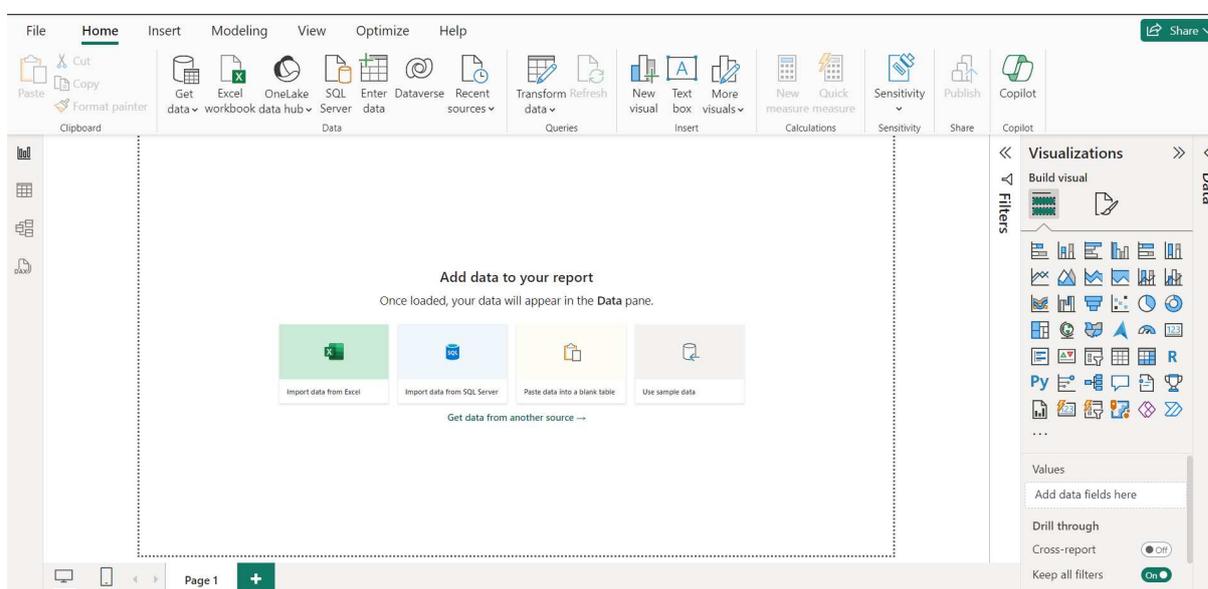
*"Dado é a personificação simplista de uma "coisa" que não nos traz nenhum sentido OU que nos gera um sentido duplo."*

No entanto os dados são a base para formar informações relevantes que é utilizado na construção do conhecimento como esquema na figura 21. Esse conhecimento é registrado e armazenado em sistemas computacionais a fim de compartilhar esse conhecimento composto do capital intelectual de diferentes colaboradores e domínios de dados (PRIMAK, Fábio Vinícius, 2008).

### 3.5.1 Ferramenta Microsoft Power BI

O *Power BI* é uma ferramenta de análise de dados desenvolvida pela *Microsoft*, que se destaca pela sua capacidade de tratamento, modelagem e análise de dados. De acordo com a (MICROSOFT, s.d.[e]), o *Power BI* é uma solução de análise de negócios que permite a visualização de dados e o compartilhamento de conhecimento gerado utilizando dados disponíveis agrupados em toda a organização, além de possibilitar a inserção de dados em aplicativos ou sites. A ferramenta se conecta a diversas fontes de dados, trazendo vida aos dados através de painéis e relatórios interativos. A interface inicial da ferramenta pode ser visualizada na figura 22

Figura 22 – Interface Power BI.



Fonte: Autor.

O *Power BI* é composto por um conjunto de serviços de *software*, aplicativos e conectores que trabalham em conjunto para transformar fontes de dados não relacionadas em informações coerentes e visualmente atraentes. Um dos destaques do *Power BI* é a utilização das fórmulas *DAX* (*Data Analysis Expressions*). A *DAX* é uma linguagem de expressão de fórmula desenvolvida pela empresa *Microsoft* utilizada na ferramenta *Power BI* entre outros produtos da empresa. Esta linguagem inclui funções, operadores e valores que possibilitam a realização de cálculos avançados e consultas em dados nas tabelas e colunas relacionadas nos modelos de dados (MICROSOFT, s.d.[e]).

Nome\_da\_medida = Função(Tabela[Coluna])

A integração dessas tecnologias permite que o *Power BI* ofereça uma plataforma robusta para a análise de dados empresariais possibilitando a modelagem de dados

complexos e a criação de visualizações dinâmicas, a ferramenta permite que os usuários a explorar dados de maneiras inovadoras e interativas. No projeto de redefinição de arquitetura de dados como o proposto, o *Power BI* desempenha um papel central. Através da conexão com diversas fontes de dados e da utilização das funcionalidades avançadas de transformação e análise de dados, ele não apenas melhora a qualidade e a acessibilidade dos dados, mas também potencializa a capacidade analítica da empresa. Assim, os relatórios interativos e os relatórios detalhados gerados pelo *Power BI* tornam-se ferramentas estratégicas para os colaboradores, promovendo uma cultura organizacional orientada á dados, o que facilita a identificação de oportunidades de melhoria e contribui para a eficácia das estratégias empresariais e apoiando decisões informadas e ágeis.

### 3.6 INDICADORES DE DESEMPENHO

Em (DOMINGUES; PEDROSA; BERNARDINO, 2021), definem-se os indicadores de desempenho (ou indicadores chave de desempenho) como sendo a representação de métricas que mensuram o rendimento empresarial de acordo com objetivos e cenários alvo de determinada organização desvendando dados através de análises estratégicas e auxiliando na tomada de decisões.

A necessidade de uma continua avaliação de desempenho através de dados que se traduzem em informações é vital para toda e qualquer organização, "o que não é medido não é gerenciado" (KAPLAN; NORTON, 1997). No âmbito da BI, os indicadores de desempenho proporcionam um panorama claro do funcionamento da organização, possibilitando a análise detalhada dos dados que alimentam painéis e relatórios no *Power BI*.

De acordo com (KIYAN, 2001), alguns dentre os diversos objetivos de realizar a medição de indicadores de desempenho é facilitar a comunicação da estratégia, instruir no entendimento dos valores e processo para alinhamento, é diagnosticas eventuais falhas, problemas e também oportunidades de melhoria. É ter controle a médio e longo prazo através de planejamento de cenários futuros comparados ao andamento do período. Identificar ações e mudanças necessárias para alcançar um resultado cada vez mais satisfatório.

### 3.7 CONTROLADORIA

De acordo com (WEISS, 2013), a controladoria foi introduzida no Brasil pelas multinacionais americanas. Elas em a finalidade de realizar um controle e planejamento a médio e longo prazo de um grupo que tem maior parte das participações de entidades. De acordo com (OLIVEIRA; PEREZ JÚNIOR; SILVA, 2009), uma controladoria eficiente e eficaz deve ser proficiente na análise e na comunicação de dados e informações

relevantes para tomada de decisão estratégicas. Além disso ser capaz de manter monitoramento e controle contínuo de diferentes atividades e desempenho das partes. Possuindo assim influência direta nas decisões gerenciais dos líderes das entidades.

De acordo com (BRUNI; GOMES, 2010), a função do profissional que trabalha na controladoria não se limita a relatar dados estáticos, ela deve demonstrar resultados das atividades operacionais de modo que influencie em decisões que direcionem a empresa para um cenário planejado.

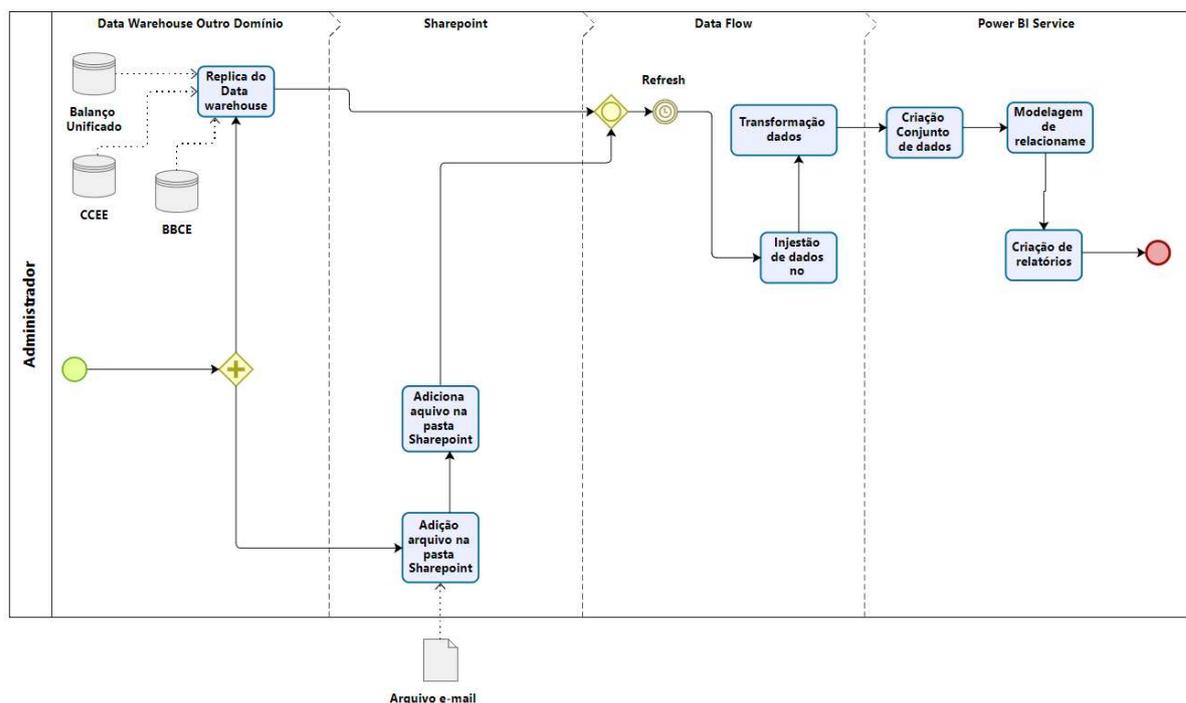
## 4 ESPECIFICAÇÕES DO PROJETO

Neste capítulo, serão discutidas especificações do projeto, assim como, a proposta de solução desenvolvida para atingir os objetivos no seu desenvolvimento fazendo uso da notação de modelagem de processos de negócio ou *Business Process Model and Notation* (BPMN), para exibir o fluxograma dos processos com intuito de documentar e de facilitar o entendimento. O objetivo deste capítulo é descrever os requisitos a serem cumpridos, fluxo de processos, atores, arquitetura, sequências, estratégias de integração entre outras informações e conceitos, os quais foram primordiais para o sucesso da solução proposta no desenvolvimento do projeto. O desenvolvimento deste trabalho é baseado na metodologia CRISP-DM, que apresenta as principais etapas e a estrutura geral utilizada para resolver o problema.

### 4.1 MAPEAMENTO DO PROCESSO/ FLUXO DOS DADOS

As informações apresentadas pela área financeira de controle de risco da empresa consomem dados de distintas fontes e domínios. O mapeamento do fluxo destes dados e de suas transformações é uma etapa vital para obter sucesso no desenvolvimento da solução proposta. Primeiramente foram identificadas as diferentes fontes de dados, as quais são extraídas e transformadas pela ferramenta *Power Query* para na sequência serem carregadas na ferramenta de *Business Intelligence*, o *Power BI*. Uma imagem do mapeamento do processo utilizando *Business Process Model and Notation* (BPMN) é apresentada na figura 23.

Figura 23 – Mapeamento processo BPMN.

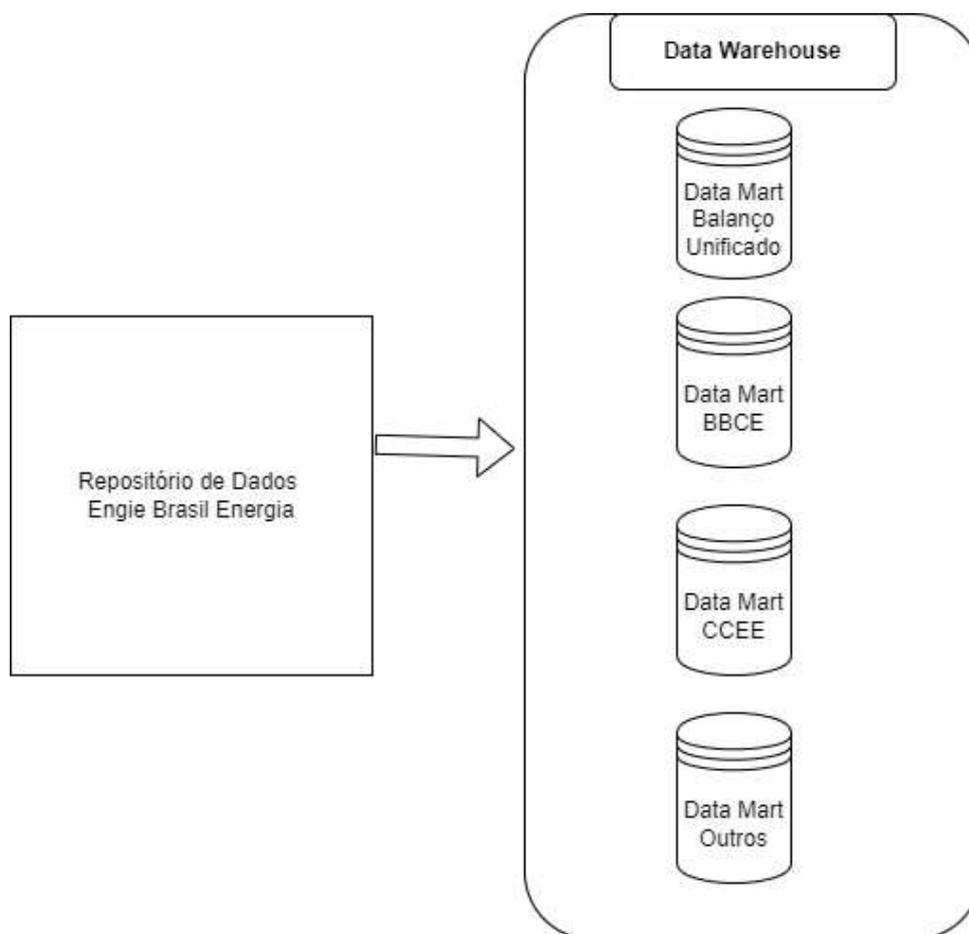


Fonte: Aluno.

Inicialmente, realiza-se um mapeamento identificando o fluxo de informações, assim como, suas fontes e origem de dados utilizadas no processo. Essas informações são utilizadas com diversos objetivos e quase totalidade delas possuem um domínio em comum que é um *Data Warehouse* central. O *Data Warehouse* consome informações de um repositório com grande volume de dados, o qual será modelada como uma "caixa preta", pelo fato deste pertence a outro domínio que a área não possui credenciais de acesso. Neste caso, a estratégia adotada é de replicar esse *Data Warehouse* da arquitetura de dados dessa origem, como mostra a figura 24.

Outra fonte de dados modelada para ser escalável é menos robusta, mas dispensa a necessidade de conhecimentos de alto nível em tecnologia, tornando-se uma utilização de fácil utilização por outros membros da equipe. Essa fonte é uma pasta nos serviços em nuvem da *Azure*, já utilizados na empresa. Nesta pasta, como mostra a figura 25, são adicionados arquivos de texto que são tabelas recebidas via correio eletrônico e adicionadas a pasta.

Figura 24 – Visão Data Warehouse Engie Brasil Energia.



Fonte: Aluno.

Figura 25 – Pasta do Sharepoint.

Name	Status	Date modified	Type	Size
abr-22	✓	04/06/2024 22:38	Microsoft Excel Com...	3 KB
jun-22	✓	04/06/2024 22:39	Microsoft Excel Com...	3 KB
mai-22	✓	04/06/2024 22:37	Microsoft Excel Com...	3 KB
mar-22	✓	04/06/2024 22:35	Microsoft Excel Com...	3 KB
v1-23	✓	04/06/2024 22:32	Microsoft Excel Com...	2 KB

Fonte: Aluno.

Todas as fontes de dados mapeadas e prontas para serem integradas, é desenvolvido um fluxo de dados, este utiliza o *Power Query Online* para se conectar as

fontes de dados desejadas. Após a integração os dados passam por transformações que os adequam para os objetivos das análises de inteligência de negócio. A etapa final desse fluxo de dados é o carregamento dos dados, na arquitetura este fluxo de dados, quando executado, cria um conjunto de dados salvo no repositório em nuvem e utilizado para alimentar a ferramenta final de visualização de relatório. O fluxo é configurado para agendar a sua execução que atualiza o conjunto de dados e consequentemente a ferramenta de BI. Esta ferramenta permite aos usuários acessar, navegar de modo que adquiram as percepções do negócio.

## 4.2 REQUISITOS FUNCIONAIS

Os requisitos funcionais definem as funcionalidades necessárias para o sucesso da solução proposta do projeto, isso é refletido na satisfação dos usuários e garantir a eficiência do fluxo de informações já mencionado. Abaixo estes requisitos elaborados conforme essas necessidades de melhorias são listados:

- Replicação do Data Warehouse: Construir uma réplica das fontes de dados contendo os *Data Marts* utilizados, permitindo a expansão para conjuntos adicionais conforme demandas futuras;
- Coleta das fontes de dados: Coletar dados através de integração das diferentes fontes com o ambiente em nuvem;
- Preparação das fontes de dados: Preparar dados oriundos da integração e através de transformações adequar os dados para os objetivos do processo de BI;
- Automatização do fluxo de dados: Estabelecer a taxa adequada de atualização para o fluxo de dados;
- Interatividade dos painéis de visualização: Construir e otimizar a interatividade dos painéis afim de facilitar a navegação e o entendimento das informações através dos visuais estratégicos.

## 4.3 REQUISITOS NÃO FUNCIONAIS

Os requisitos não funcionais são definidos com o conceito de analisar qualitativamente as capacidades do sistema visando uma operação eficiente, segura e satisfatórias do mesmo. Leva-se em conta também a manutenção de desempenho ao longo do tempo de vida do projeto.

- Desempenho: O sistema deve ser eficiente, possuir tempos de atualização adequados para uma visão atualizada dos dados, processar e acessar dados em um tempo satisfatório para usuários finais e administradores;

- **Segurança:** Garantir a integridade e confiabilidade dos dados sensíveis ao longo de todo o fluxo de informações com adequado controle de acesso;
- **Escalabilidade:** A arquitetura deve ser escalável prevendo suportar o aumento futuro de volume de dados;
- **Usabilidade:** O sistema deve ser intuitivo para os usuários finais, requerendo baixa manutenção e sendo facilmente editável e acessível para os administradores;
- **Confiabilidade:** O sistema deve manter a integridade dos dados, disponibilizando acesso ao conjunto de sistemas integrados a qualquer momento via rede.

#### 4.4 PROPOSTA DE SOLUÇÃO

A proposta de solução pode ser segmentada em duas principais partes, a implantação de uma nova arquitetura de dados e remodelagem do sistema de *BI*. Abaixo serão listados os processos AS-IS e TO-BE de maneira a esclarecer os desafios e a solução para resolvê-los.

##### AS-IS:

- Coleta e integração de dados desorganizados e pouco eficientes de fontes muitas vezes redundantes;
- Arquitetura vulnerável a falhas e inconsistências;
- Necessidade de uma grande quantidade de horas de trabalho manual;
- Processo ETL e organização de dados inadequadas para interatividade desejada;
- Painéis de visuais pouco intuitivos e interativos;
- Ausência de rotina de atualização automática do processo de ETL dos dados usados nos painéis.

##### TO-BE:

- Reestruturar a arquitetura de dados;
- Integração de fonte de dados organizadas sem redundância;
- Automatizar processos de atualização dos dados agendados por rotina;
- Remodelar processo de transformação dos dados para integrar com interatividade dos painéis;

- Remodelar os painéis e visuais para aumentar a interatividade e facilitar o entendimento do usuário final;

Na implantação de uma nova arquitetura de dados, propõe-se a replicação de um repositório pertencente de outra entidade, com os conjuntos de dados a serem utilizados. Este repositório possui informações do grupo ENGIE no Brasil, alguns deles já transformados contendo diversas tabelas e outros sendo apenas uma tabela única. Esta arquitetura é a origem de quase a totalidade das fontes de dados utilizadas.

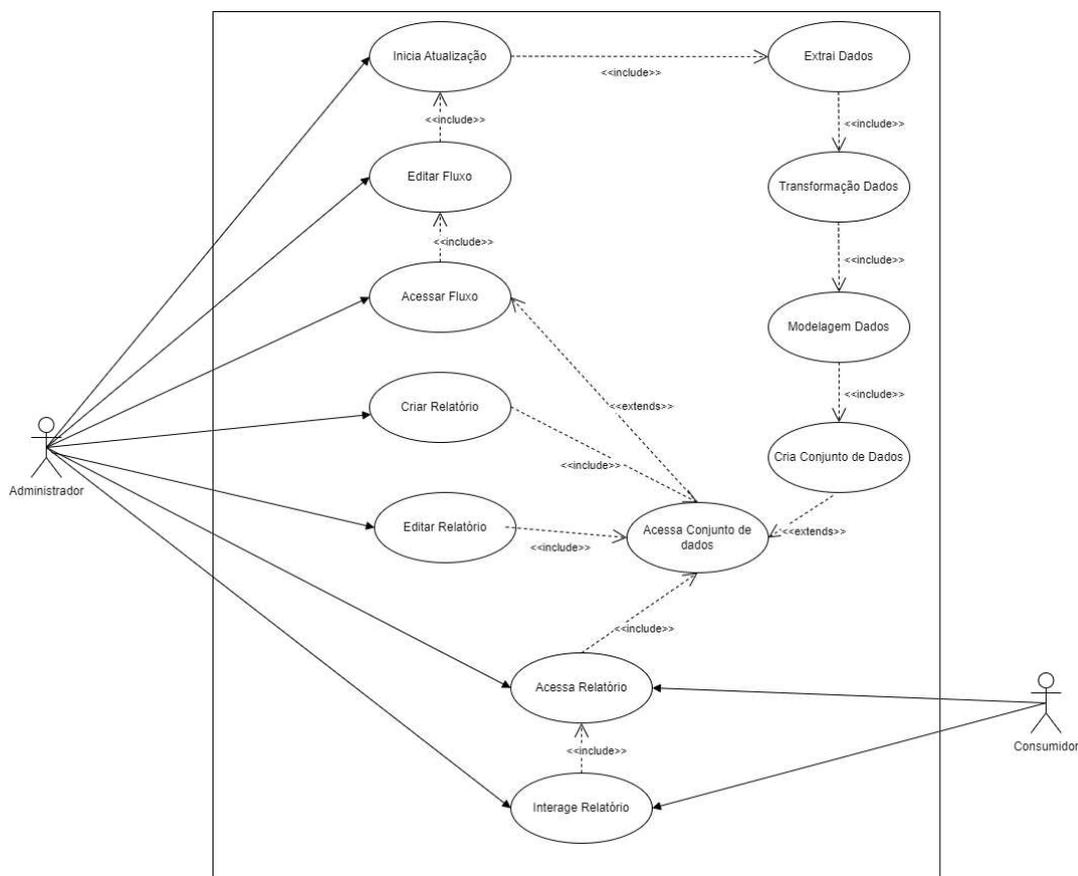
É proposto a replicação para gerenciar mudanças e garantir acesso livre a equipe de controle de risco. Após o mapeamento e a replicação desse repositório, no formato de um *Data Warehouse* contendo os *Data Marts* selecionados. É integrado também uma pasta, com armazenamento em nuvem, do *Sharepoint* para ingestão destes dados no fluxo presente na arquitetura. Essa integração adicional é realizada através da instalação e integração de um *gateway* de dados.

Os dados são integrados com o fluxo de dados no serviço em nuvem, passam por transformações e são carregados em um conjunto de dados também inseridos no mesmo serviço em nuvem da *Microsoft*. Esse processo é denominado ETL e é agendado, no fluxo de dados, através de configurações disponibilizadas na ferramenta, atualizações automáticas deste processo. Deste modo é possível criar conjunto de dados sempre baseados em dados recentes para o processo de BI.

Sequencialmente, os dados processados são modelados por relacionamentos constituindo modelo de dados que serão carregados no relatório da área. Como já explicado, este conjunto de dados é criado no sistema de nuvem da Microsoft e representa um Data Mart ou Data Warehouse, ele fica disponível para criação de outros relatórios por outras equipes com as credenciais necessárias.

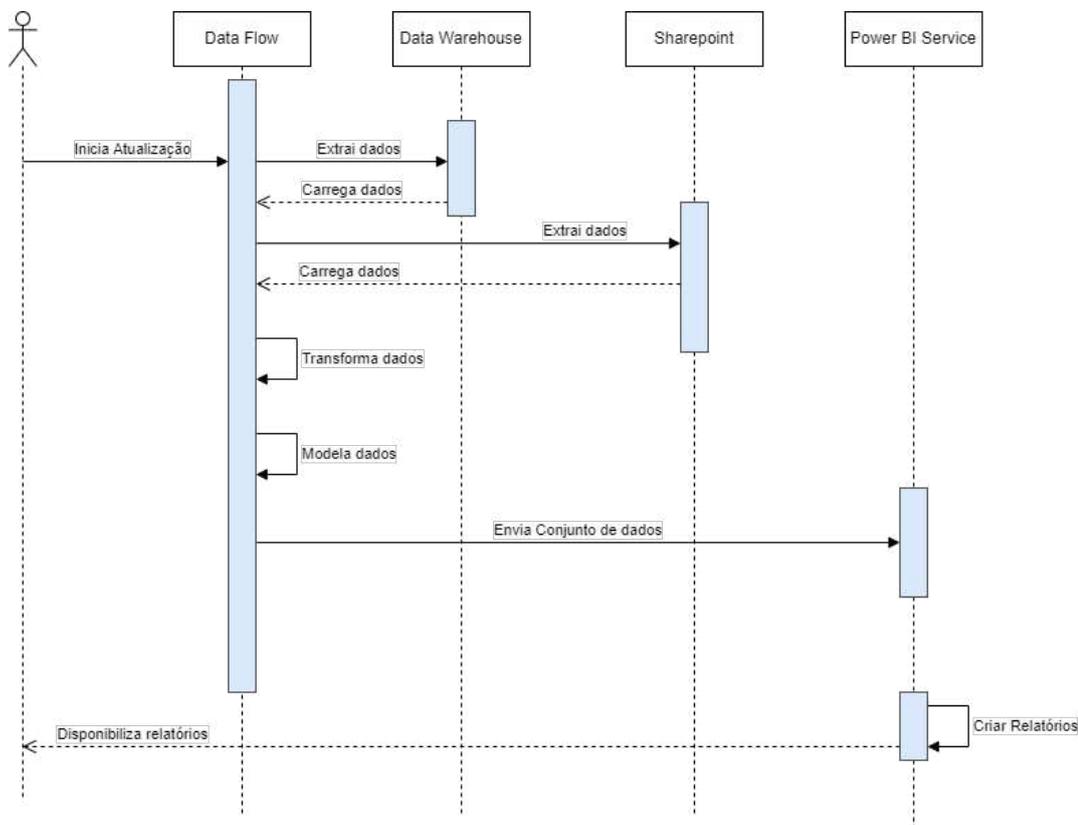
Na segunda parte são criados e modelados os relatórios que consomem o conjunto de dados na arquitetura de dados. Ao final esse conjunto de dados, no formato adequado, são carregados para serem utilizados em análises estratégicas de *Business Intelligence*. Estas ficam disponíveis a qualquer momento, de forma visual através da ferramenta para consulta de diretores, líderes e colaboradores com devido credenciamento. Estes usuários finais interagirão com a ferramenta navegando entre as páginas e selecionando filtros no *Power BI* como mostra a figura (Caso de uso) para fácil navegação entre versões, entidades, datas entre outras categorias disponíveis dependendo da necessidade específica para um grau de transmissão de conhecimento satisfatório através dos relatórios interativos.

Figura 26 – Diagrama UML Casos de Uso da proposta de solução do projeto.



Fonte: Aluno.

Figura 27 – Diagrama UML de Sequência da proposta de solução do projeto.



Fonte: Aluno.

## 5 DESENVOLVIMENTO

Nos capítulos anteriores foram apresentadas diversas informações sobre o projeto, essas serviram de base para o desenvolvimento da solução proposta. Foram explicitadas as problemáticas e contextos gerais sobre ferramentas utilizadas, assim como, os fundamentos conceituais base para a aplicação. Foi descrito o panorama geral do setor de energético brasileiro e da área financeira de controle de risco ligados empresa Engie Brasil Participações, na qual o projeto foi realizado. Também foram apresentados o escopo do projeto, diagramas, requisitos e mapeamento da solução proposta.

Neste capítulo são detalhadas as etapas executadas para solução do problema desenvolvido através metodologia escolhida. O desenvolvimento desta solução pode ser separado na parte da arquitetura que define o fluxo de dados e da ferramenta de BI que é o objetivo final do projeto. A aplicação deste desenvolvimento é apresentada na prática para sequencialmente avaliar seus resultados. Os dados utilizados no projeto não refletem a situação atual da empresa, são dados fictícios, construídos com estrutura e características de modo a simular a situação de uma empresa do mesmo setor.

### 5.1 ETAPAS DESENVOLVIMENTO

Nesta seção, são abordados o desenvolvimento fragmentado de cada etapa da construção da solução utilizando os conhecimentos já apresentados e a metodologia proposta.

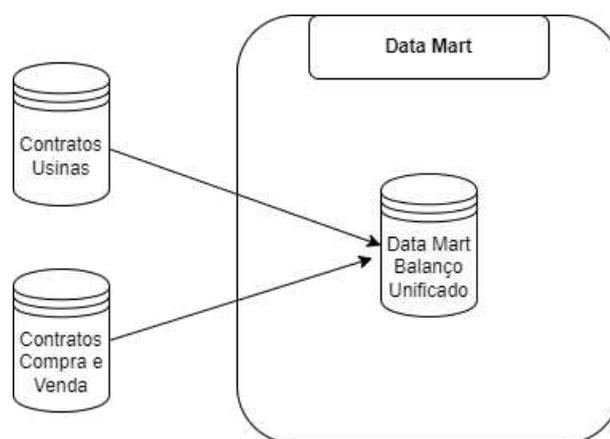
#### 5.1.1 *Business Understanding*

Nesta fase inicial de análise, foi realizado um mapeamento abrangente do atual fluxo de dados presente na arquitetura de dados utilizadas na área financeira de controle de riscos da empresa do setor energético Engie no Brasil. Esta etapa é a base de todas as outras, por esta razão foi investido um tempo considerável do projeto em seu desenvolvimento. Pelo conjunto de ferramentas inadequados e pelas fontes de dados pertencerem a outras áreas, levou-se um tempo para entender o negócio para elucidar os objetivos, as necessidades alinhando todo esse processo mapeamento entre as áreas envolvidas no projeto.

O objetivo principal do projeto é desenvolver um relatório com uma visão de negócio interativo, possibilitando consultas a qualquer momento para embasar decisões estratégicas. Essa visão de negócio é baseada em dados de contratos de compra e venda de energia realizados, podendo ser através plataforma BBCE, de forma bilateral ou para o mercado regulado. Também de preços da energia do órgão governamental

CCEE, preços de energia estimados da empresa, volume produzidos por usinas, valores planejados de orçamento, cenários futuros, indicadores de desempenho entre outros. O exemplo do esquema da origem que foram o *Data Mart* Balanço Unificado é apresentada na figura 28

Figura 28 – Construção Data Mart.



Fonte: Aluno.

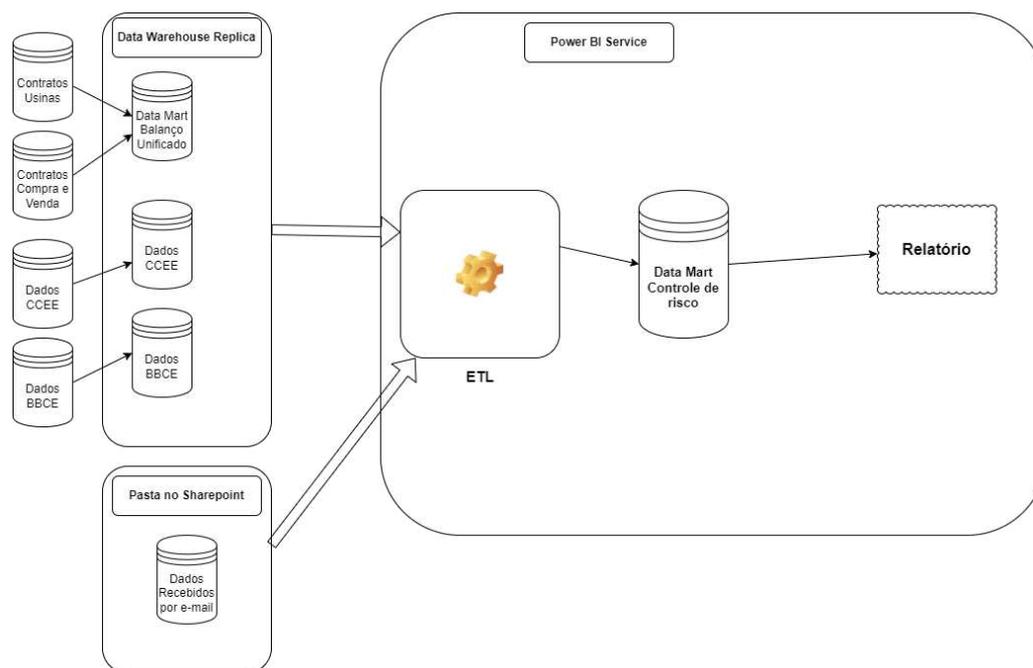
Na ótica dos integrantes da área o objetivo é automatizar esse fluxo de forma agendada de forma que diminua ou elimine a necessidade de interação humana. Desse modo mantendo ou aumentando a acurácia dos dados, otimizando a eficiência do processo de BI e aumentando o poder de transmitir conhecimentos estratégicos. Na ótica dos diretores e líderes, os quais são os clientes finais desse projeto, o objetivo é utilizar a ferramenta de forma mais eficiente e intuitiva, o que se deve á reestruturação da arquitetura de dados. O aumento da granularidade e agendamento de versões comparado com o processo antigo é um exemplo de melhoria que traz valor a empresa através análises de dados mais recentes. Na entrega final do projeto sendo os relatórios interativos da ferramenta *Power BI* serão acessados a qualquer momento de modo a criar compreensão e entendimento dos dados de maneira intuitiva, alcançando objetivos de inteligência de negócio.

### 5.1.2 Data understanding

Após o entendimento dos objetivos e das necessidades do negócio, na segunda etapa do CRIS-DM são mapeadas as diferentes fontes de dados, suas origens e seus domínios, ou o setor da empresa á quem este conjunto de dados pertence. Ao mapear as fontes de dados levando em conta suas origens e seus domínios entendeu-se que a grande maioria delas têm origem em comum. Sendo origem em comum um *Data Warehouse* da empresa *Amazon*, o *AWS Redshift* sob o domínio de outra entidade do grupo ENGIE, o Engie Brasil Energia.

A outra fonte de dados é recebida por correio eletrônico oriunda da área de planejamento e análise financeira, essa foi modelada através de arquivos em formato de texto adicionados manualmente a uma pasta do sistema de armazenamento em nuvem da *Microsoft* chamado *One drive*, e contido dentro dos serviços em nuvem da *Azure*. Um esquema do entendimento dos dados é apresentado na figura 29

Figura 29 – Esquema Entendimento dos dados.



Fonte: Aluno.

Iniciando entendimento dos dados pelas fontes de dados mais significativa para o projeto, é realizada uma análise do *Data Warehouse* com objetivo de entender os conjuntos de dados contidos nos *Data Marts* e as informações as quais eles representam. Desse modo, a fase seguinte de preparação dos dados para análises, seja mais assertiva.

- Conjunto de dados do *Data Mart* Balanço Unificado : Usinas e Contratos: Este *Data Mart* é utilizado pelo setor de gerenciamento de portfólio. Os dados desta fonte originam-se de dados adquiridos, processados e armazenados das usinas e sistemas de atores do grupo. Estes formam informações das métricas de volume das usinas e informações de contratos das compras e vendas que da mesma forma são coletados, processados e armazenados formando informações de preços e volume. Esses conjuntos de dados relacionados entre seus bancos de dados, formando informações sobre tipo tecnologia, tipo de operação do mercado livre ou regulado, nome da entidade, data mensais, versões, data das versões entre outros como mostra a tabela 30:

Figura 30 – Tabela de dados Data Mart do Balanço EBE.

Entity	Activity	Type	Techno	Agrupamento	Date	Volume	Price	View_date	Version
Geramamore				Open position	01/01/2023	5.83822691939082		0 jan/22-2	79999401
EBE	Generation	Regulated	Solar	Solar regulated	01/01/2023	68.5046847910423		0 jan/22-2	79999401
EBE	Generation	Regulated	Biomass	Biomass regulated	01/01/2023	18.830855758642		0 jan/22-2	79999401
EBE	Generation	Regulated	Thermal	Thermal	01/01/2023	31.0621798988805		0 jan/22-2	79999401
EBE	Generation	Regulated	Hydro	Hydro regulated	01/01/2023	564.763699936146		0 jan/22-2	79999401
EBE	Generation	Free market	Hydro	Hydro free market	01/01/2023	1382.85236963631		0 jan/22-2	79999401
EBE	Generation	Free market	Wind	Wind free market	01/01/2023	116.366727397557		0 jan/22-2	79999401
EBE	Generation	Free market	Solar	Solar free market	01/01/2023	-6.08999719985229E-05		0 jan/22-2	79999401
EBE	Generation	Free market	Biomass	Biomass free market	01/01/2023	-26.4249650632839		0 jan/22-2	79999401
EBE	Purchases	Free market	Market	OTC fwd purchases	01/01/2023	0		0 jan/22-2	79999401
EBE				Open position	01/01/2023	615.162982042782		0 jan/22-2	79999401
Geramamore	Sales			OTC fwd sales	01/01/2023	184.306251824358		0 jan/22-2	79999401
Jirau	Generation	Regulated	Hydro	Hydro regulated	01/01/2023	280.101846826923		0 jan/22-2	79999401
Jirau	Generation	Free market	Hydro	Hydro free market	01/01/2023	112.887766244729		0 jan/22-2	79999401
Jirau				Open position	01/01/2023	19.5715665807065		0 jan/22-2	79999401
EBE	Generation	Regulated	Wind	Wind regulated	01/01/2023	138.207944959993		0 jan/22-2	79999401
EBE	Purchases	Free market	Market	LT purchases	01/01/2023	213.158884675748	250.317833464756	jan/22-2	79999401
Jirau	Purchases	Free market	Market	LT purchases	01/01/2023	51.5823743739943	157.358028729534	jan/22-2	79999401
Geramamore	Purchases			SPPA purchase (Engie)	01/01/2023	196.760744095067	250.38785141394	jan/22-2	79999401
Geramamore	Purchases			SPPA purchase (Mitsui)	01/01/2023	49.7266128227018	133.128689546388	jan/22-2	79999401
Geramamore	Sales			BtoB sales	01/01/2023	11.1078646113652	141.008450584063	jan/22-2	79999401

Fonte: Aluno.

Nesta fase também é mapeado a frequência de geração de novas versões desse conjunto de dados, essa atualização é de duas vezes ao dia, sendo uma atualização as duas horas e outra sendo as onze horas. Cada uma das versões pode ser entendida como uma captura, no momento da data e hora da versão, do passado e futuro relativos dos preços e volumes de compra, venda e geração de energia por entidades do grupo Engie no Brasil. Os dados são categorizados neste formato de versão com visões do concretizadas e estimadas em forma de datas no passado e no futuro. Desse *Data Mart* são utilizadas duas tabelas, uma fato contendo as informações gerais e uma de versionamento. Estes dados já estão previamente tratados de forma eficiente e com segmentações que serviram de fonte de dados para o projeto.

- Tabela do *Data Mart* BBCE: Esta fonte de dados pertence ao setor de *Trading* do grupo Engie no Brasil e pode ser visualizada uma amostra na figura 31. Diferente do anterior, esta fonte é um banco de dados que armazena informações sobre operações realizadas, disponibilizados pela plataforma do Balcão Brasileiro de Comercialização de Energia (BBCE). Este banco de dados é composto por operações de compra e venda realizadas. Cada operação é identificada pela chave primária ID único da operação que possui informações Como: data e hora de conclusão, descrição com informações de intervalos de mês e ano de entrega do produto, tipo do produto. Além disso possui métricas de preço e volume refe-

reenciando cada transação concluída. Todas essas informações contidas em uma tabela combinada única necessita do processo de Extração, Transformação e Carregamento (ETL) para se adequar aos uso final.

Figura 31 – Tabela de dados da plataforma BBCE.

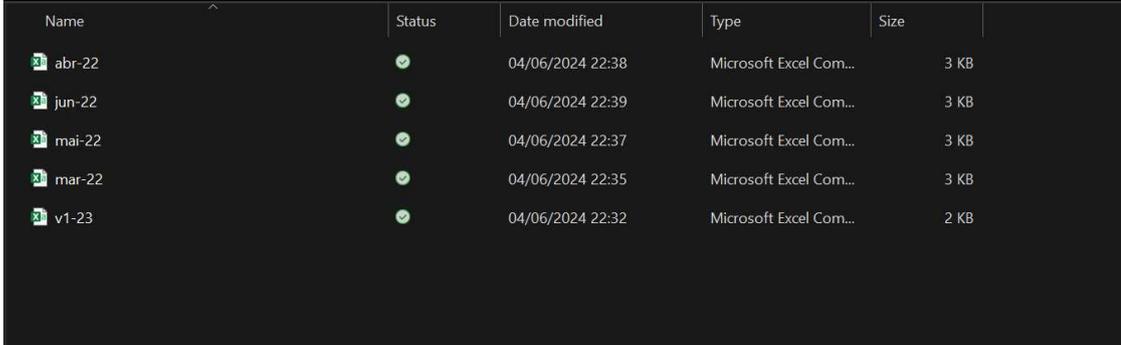
Preço	Volume	Tipo de Produto	Start Date	Final	Cód. do Produto	Nr. da Operação	Data Operação
93.5504038095998	1.05948680754	MEN	segunda-feira, 1 de julho de 2024		1340	990698	quinta-feira, 4 de abril de 2024
83.2379535037683	5.41091171011	MEN	segunda-feira, 1 de julho de 2024		1344	989388	terça-feira, 2 de abril de 2024
93.1027307616727	4.87263684741	MEN	segunda-feira, 1 de julho de 2024		1360	989374	terça-feira, 2 de abril de 2024
83.6189160659625	4.99612714659	MEN	segunda-feira, 1 de julho de 2024		1684	989332	terça-feira, 2 de abril de 2024
72.8167947050953	1.08855170126	MEN	segunda-feira, 1 de julho de 2024		2644	990857	quinta-feira, 4 de abril de 2024
73.8942380610302	1.84558407243	MEN	segunda-feira, 1 de julho de 2024		2922	992488	terça-feira, 9 de abril de 2024
55.2582479153285	1.93306683493	MEN	segunda-feira, 1 de julho de 2024		2938	992489	terça-feira, 9 de abril de 2024
81.1993247419759	0.95488857540	MEN	segunda-feira, 1 de julho de 2024		3044	992244	terça-feira, 9 de abril de 2024
76.6427842308058	3.01658644093	MEN	segunda-feira, 1 de julho de 2024		3176	990053	quarta-feira, 3 de abril de 2024
62.222130043382	0.64495538679	MEN	segunda-feira, 1 de julho de 2024		3254	992856	quinta-feira, 11 de abril de 2024
75.2373062692215	0.74650413757	MEN	segunda-feira, 1 de julho de 2024		3430	993095	quinta-feira, 11 de abril de 2024
82.5041143873818	0.90444640240	MEN	segunda-feira, 1 de julho de 2024		3462	990120	quarta-feira, 3 de abril de 2024
65.3933043169947	2.14694982536	MEN	segunda-feira, 1 de julho de 2024		3500	993969	quinta-feira, 18 de abril de 2024
75.4623552073273	4.78994655938	MEN	segunda-feira, 1 de julho de 2024		3506	990177	quarta-feira, 3 de abril de 2024
58.936661932221	1.81321046001	MEN	segunda-feira, 1 de julho de 2024		3556	993978	quinta-feira, 18 de abril de 2024
73.7228276736221	1.02912094702	MEN	segunda-feira, 1 de julho de 2024		3614	993985	quinta-feira, 18 de abril de 2024
70.9318085677585	2.16536237571	MEN	segunda-feira, 1 de julho de 2024		3694	991765	segunda-feira, 8 de abril de 2024
80.1188782851841	1.09649255835	MEN	segunda-feira, 1 de julho de 2024		3740	992176	terça-feira, 9 de abril de 2024
58.2233695305612	197.185213416	MEN	segunda-feira, 1 de julho de 2024		3790	994651	terça-feira, 23 de abril de 2024
72.0226499985836	4.98086215799	MEN	segunda-feira, 1 de julho de 2024		3874	995520	segunda-feira, 29 de abril de 2024

Fonte: Aluno.

- Tabela do *Data Mart* CCEE: Analogamente ao *Data Mart* da BBCE, este captura os preços históricos de liquidação da energia brasileira da Câmara Comercializadora de Energia Elétrica (CCEE). Estes dados formam uma tabela com atributo "região" que é subdividido por regiões as quais, por consenso do grupo no Brasil é utilizada como oficial a região sudeste (SE) por ser a que mais reflete a situação de todo território nacional. Também é composta por preços históricos e datas do passado com granularidade de horas.
- Base de dados preços futuros: Esta base de dados está localizada no sistema em nuvem do *Sharepoint* e armazena em um repositório no *Microsoft Azure*, amplamente utilizado o *One Drive* sincroniza arquivos comuns entre um ou mais *desktops* que compartilham acesso a pasta através de um *Gateway*. Este é configurada para persistir uma cópia dos arquivos no banco de dados sempre que uma mudança for salva. A solução proposta para essa fonte é inserir arquivos de texto padronizados em uma pasta no *Sharepoint* e nomeá-los em um padrão acordados de antemão com a equipe que fornece os arquivos por via correio eletrônico. Estas pastas são persistidas no banco de dados na nuvem. Essa estrutura não é a mais adequada do ponto de vista de arquitetura de dados, entretanto ela formaliza um método comumente utilizado para manter uma opção

de adição de análises pontuais e adição de análises por administradores não especializados em TI.

Figura 32 – Pasta Sharepoint.



Name	Status	Date modified	Type	Size
abr-22	✓	04/06/2024 22:38	Microsoft Excel Com...	3 KB
jun-22	✓	04/06/2024 22:39	Microsoft Excel Com...	3 KB
mai-22	✓	04/06/2024 22:37	Microsoft Excel Com...	3 KB
mar-22	✓	04/06/2024 22:35	Microsoft Excel Com...	3 KB
v1-23	✓	04/06/2024 22:32	Microsoft Excel Com...	2 KB

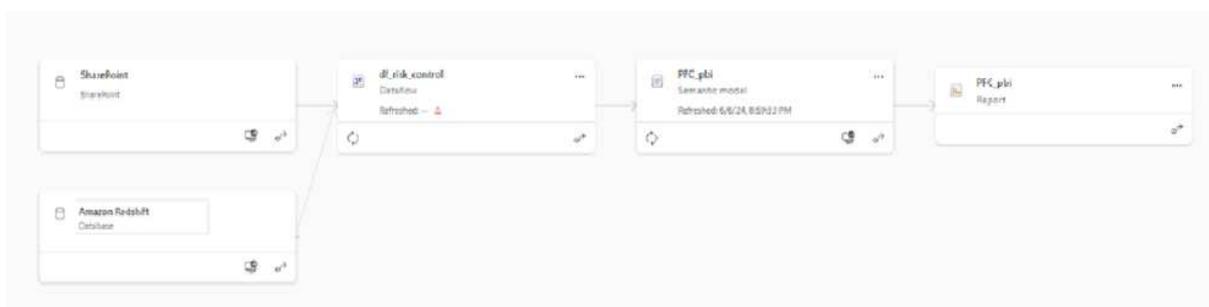
Fonte: Aluno.

Estes dados são compostos por atributos de data do produto mensal, preços futuros e a *View date* que, analogamente a construção da versão no *Data Mart* da EBE, representa uma foto da visão da empresa naquela data. Este campo de *View date* previamente acordado será o padrão de nome dos arquivos no formato de texto dentro de uma pasta na nuvem. O formato deste atributo é das três primeiras letras do mês em português, seguido por uma caractere de "-" e os últimos dois números do ano e é alinhado com a equipe que fornece estes dados a qual se comprometeu a adicionar os arquivos na pasta mensalmente resultando na pasta da figura 32.

### 5.1.3 Data preparation

Esta fase começa com a extração dos dados identificados nas fases anteriores e tem como objetivo executar o processo de ETL preparando os dados para a modelagem. Esta fase é realizada em três ambientes. Os componentes da arquitetura podem ser visualizados na figura 33

Figura 33 – Fluxo de dados.

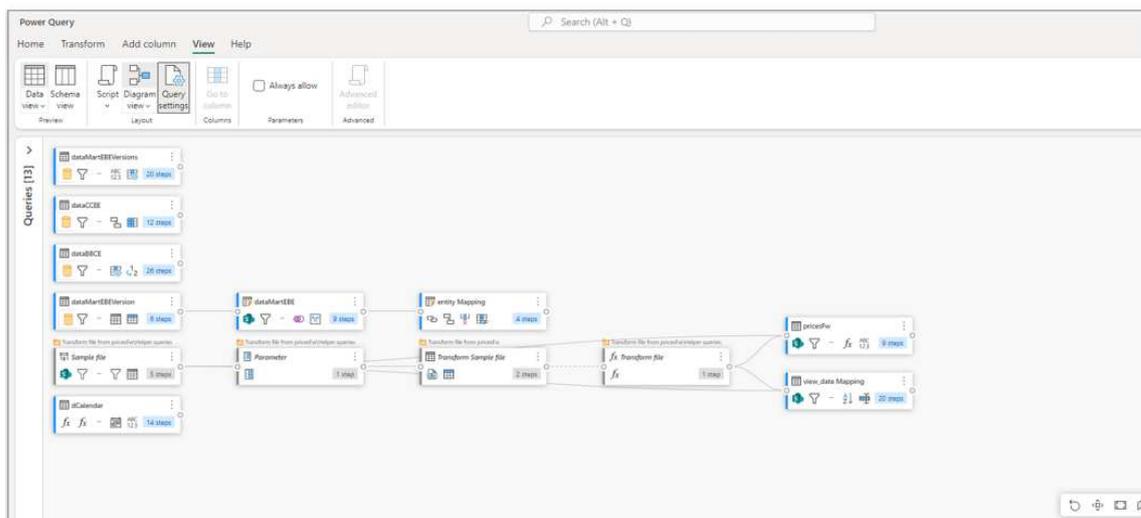


Fonte: Aluno.

A etapa inicial, realizada pela equipe de TI parceira no ambiente de dados de outra entidade, é a de replicação do *Data Warehouse* de propriedade da Engie Brasil Energia contendo os conjuntos de dados mapeados. Essa é realizada com o objetivo de gerenciar as mudanças nas estruturas dos conjuntos utilizados diminuindo a vulnerabilidade a falhas com mudanças realizadas por outras equipes sem previa comunicação.

O segundo ambiente onde essa fase acontece é o *Data Flow*, essa ferramenta é responsável pelo processo de ETL (Extract, Transform and Load), ele é responsável por definir e agendar o fluxo dos dados na arquitetura. Utilizando o fluxo de dados é possível realizar todo processo de ETL, o qual gera o conjunto de dados e o relatório, no sistema em nuvem. O processo de ETL centraliza os dados consistentes de várias fontes em um único repositório para análise de forma rápida e eficiente. Na figura 34 são apresentados detalhes sobre cada etapa deste processo realizado dentro espaço de trabalho da área de controle de risco no serviço da ferramenta *Power BI*.

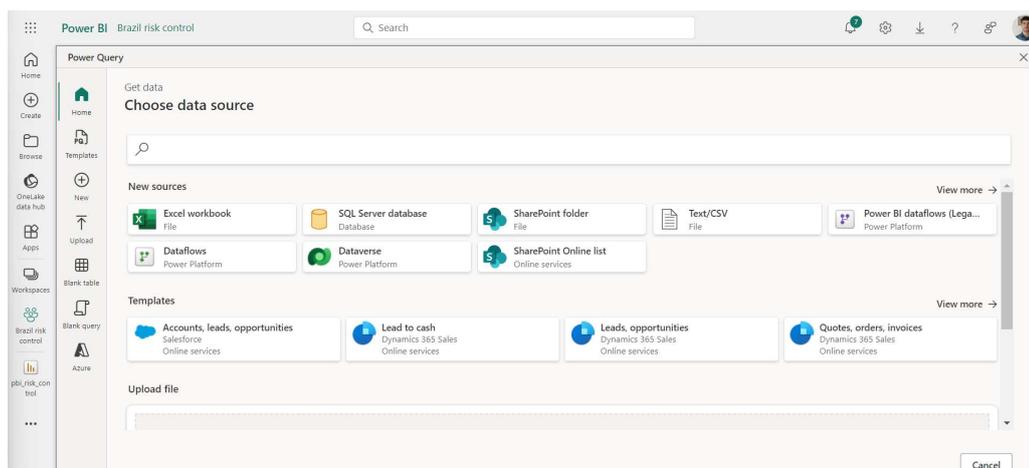
Figura 34 – Fluxo do ETL na ferramenta Power Query Online.



Fonte: Aluno.

*Extract:* A etapa de extração dos dados ocorre por duas diferentes integrações com fonte de dados, o *Data Warehouse* e pasta do *Sharepoint*. Para realização da fase do tipo de integração de banco de dados é realizada através de um gateway da própria ferramenta de *Business Intelligence*. As interfaces de configuração da integração com as fontes são apresentadas nas figuras 35, 36 e 37.

Figura 35 – Painel inicial da extração de dados.

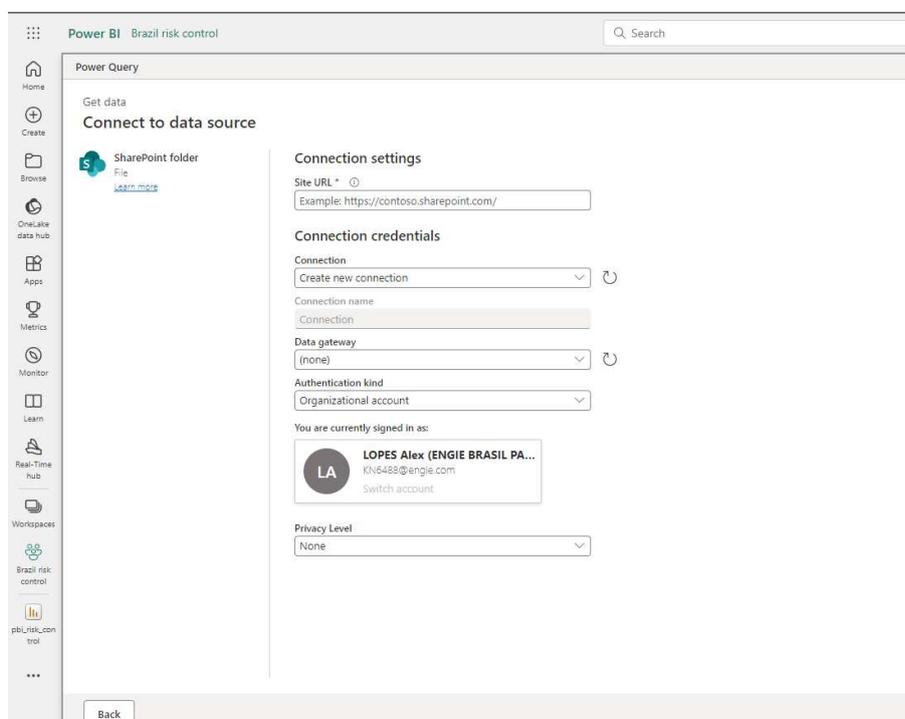


Fonte: Aluno.

O *Power Query M* é a linguagem de programação baseada em *Sequence Query Language* (SQL) criada para realizar o processo de ETL. Esta é utilizada nesta etapa

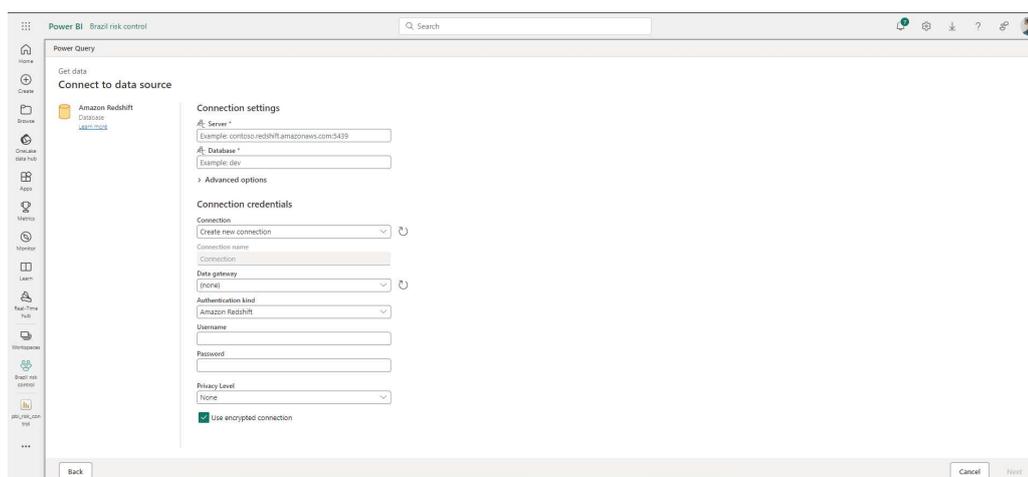
para integrar as fontes de dados ao fluxo na arquitetura. Utilizando identidades credenciadas é possível configurar a criação desse vínculo entre o fluxo e a fonte.

Figura 36 – Extração pasta Sharepoint.



Fonte: Aluno.

Figura 37 – Extração DW Amazon Redshift.



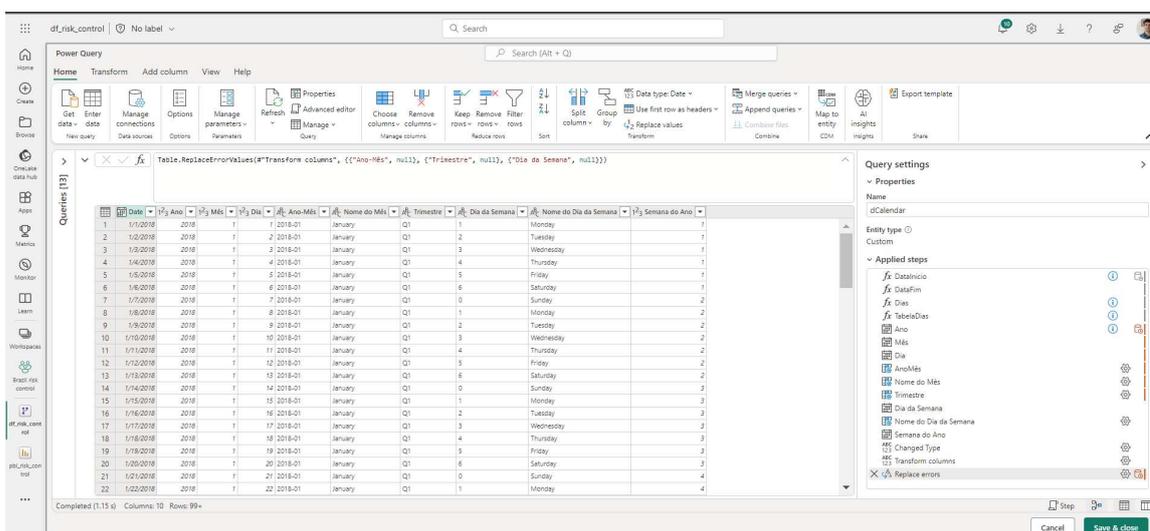
Fonte: Aluno.

*Transform:* O objetivo da etapa de transformação é processar dados em um formato adequado para análises, esse objetivo é alcançado através de processos de

remoção de inconsistências, transformações lógicas manipulando dados e enriquecimento de informações através de adição ou derivação de um ou mais banco de dados. Esse processo ocorre no visual de duas maneiras, através a linguagem Power Query ou Dax dependendo do uso e do conhecimento em cada uma delas. Sendo a linguagem Dax responsável apenas pela transformação, mais voltada para a interação com os visuais dentro dos relatórios.

**Power Query Editor:** Essa ferramenta é separa por etapas, disponibilizando resultados intermediários e podendo conectar a esses resultados como mostrada na figura 38 da interface do Power Query. As primeiras etapas são de navegação até o local dos arquivos das fontes de dados, estes serão disponibilizados os códigos nos apêndices ao final do documento.

Figura 38 – Interface Power Query Online.



Fonte: Aluno.

Tabela dados: É apresentado abaixo o desenvolvimento da transformação de dados no *Power Query* separada por tabelas.

- **Balanco Unificado de Versões:** O processo de transformação da tabela que armazena informações das versões do Balanço de Usinas e Contratos. Após navegação e rotulagem de dados é criado um filtro de dias selecionando apenas datas de versões entre o dia 03 e 05 e entre os dias 13 a 15. A justificativa da seleção destes parâmetros é que a taxa de atualização desejada é entre esses dias.

Após essa etapa é adicionada uma coluna para registrar se é da primeira ou segunda versão, para depois agrupar calculando a hora máxima com objetivo de filtrar baseado na versão mais recente. Utilizando a mesma heurística, é executado um conjunto de etapas que agrupa com cálculos e enriquece dados

através da adição de colunas utilizadas em filtros como mostra a parte do fluxo na figura 39.

Figura 39 – Transformação de dados da tabela de versões do Balanço Unificado.



Fonte: Aluno.

Essas etapas reduzem a quantidade de armazenamento de versões não utilizadas consequentemente aumentando a eficiência do sistema. São utilizadas na tabela Balanço Unificado de Usinas e Contratos para não armazenar versões não utilizadas.

- Balanço Unificado de Usinas e Contratos: O processo realizado nesta tabela além da navegação, é a rotulagem dos dados para tipos de texto, número decimal e data. Após esse passo é realizada a operação de mesclar tabelas utilizando a chave primária que é um campo único da tabela Balanço Unificado de versões. O objetivo desse passo é importar as datas da versões de modo que filtre apenas as versões existentes na tabela de versões através da configuração de mesclagem *Right Outer* como demonstrado na figura 40. Essa configuração traz apenas os valores da chave primária de versão que existe na segunda tabela, a Balanço Unificado de versões, para depois expandir informações com as datas das versões.

Figura 40 – União de tabelas (Merge).

✕

## Merge

Select a table and matching columns to create a merged table.

datamartEBE 📄

Entity	Activity	Type	Techno	Agrupamento	Date	Volume	Price	Version
EBE	Generation	Regulated	Hydro	Hydro regulated	01/01/2023	1145.71257	0	79999
EBE	Generation	Free market	Hydro	Hydro free market	01/01/2023	754.8043466	0	79999
EBE	Purchases	OTC fwd purchases	Market	OTC fwd purchases	01/01/2023	464.0942916	283.3062734	79999
EBE	Generation	Regulated	Wind	Wind regulated	01/01/2023	240.2028399	0	79999

datamartEBE\_versions ▾

Quinzena	MesAno	Date	Version	Hora	Description	View_date
1	jan/22	05/01/2022	79999784	22	version	jan/22-1
2	jan/22	15/01/2022	79999401	14	version	jan/22-2
1	fev/22	05/02/2022	79999416	20	version	fev/22-1
2	fev/22	15/02/2022	79999831	20	version	fev/22-2
1	mar/22	05/03/2022	79999445	23	version	mar/22-1

Join Kind

Right Outer (all from second, matching from first) ▾

Use fuzzy matching to perform the merge

▸ Fuzzy matching options

✓ The selection matches 5 of 30 rows from the second table.

OK
Cancel

Fonte: Aluno.

Resultando na tabela 41:

Figura 41 – Tabela final Balanço Unificado.

Entity	Activity	Type	Techno	Agrupamento	Date	Volume	Price	View_date	Version
Geramamore				Open position	01/01/2023	5.83822691939082	0	jan/22-2	79999401
EBE	Generation	Regulated	Solar	Solar regulated	01/01/2023	68.5046847910423	0	jan/22-2	79999401
EBE	Generation	Regulated	Biomass	Biomass regulated	01/01/2023	18.830855758642	0	jan/22-2	79999401
EBE	Generation	Regulated	Thermal	Thermal	01/01/2023	31.0621798988805	0	jan/22-2	79999401
EBE	Generation	Regulated	Hydro	Hydro regulated	01/01/2023	564.763699936146	0	jan/22-2	79999401
EBE	Generation	Free market	Hydro	Hydro free market	01/01/2023	1382.85236963631	0	jan/22-2	79999401
EBE	Generation	Free market	Wind	Wind free market	01/01/2023	116.366727397557	0	jan/22-2	79999401
EBE	Generation	Free market	Solar	Solar free market	01/01/2023	-6.08999719985229E-05	0	jan/22-2	79999401
EBE	Generation	Free market	Biomass	Biomass free market	01/01/2023	-26.4249650632839	0	jan/22-2	79999401
EBE	Purchases	Free market	Market	OTC fwd purchases	01/01/2023	0	0	jan/22-2	79999401
EBE				Open position	01/01/2023	615.162982042782	0	jan/22-2	79999401
Geramamore	Sales			OTC fwd sales	01/01/2023	184.306251824358	0	jan/22-2	79999401
Jirau	Generation	Regulated	Hydro	Hydro regulated	01/01/2023	280.101846826923	0	jan/22-2	79999401
Jirau	Generation	Free market	Hydro	Hydro free market	01/01/2023	112.887766244729	0	jan/22-2	79999401
Jirau				Open position	01/01/2023	19.5715665807065	0	jan/22-2	79999401
EBE	Generation	Regulated	Wind	Wind regulated	01/01/2023	138.207944959993	0	jan/22-2	79999401
EBE	Purchases	Free market	Market	LT purchases	01/01/2023	213.158884675748	250.317833464756	jan/22-2	79999401
Jirau	Purchases	Free market	Market	LT purchases	01/01/2023	51.5823743739943	157.358028729534	jan/22-2	79999401
Geramamore	Purchases			SPPA purchase (Engie)	01/01/2023	196.760744095067	250.38785141394	jan/22-2	79999401
Geramamore	Purchases			SPPA purchase (Mitsui)	01/01/2023	49.7266128227018	133.128689546388	jan/22-2	79999401
Geramamore	Sales			BtoB sales	01/01/2023	11.1078646113652	141.008450584063	jan/22-2	79999401

Fonte: Aluno.

- Preços futuros: O processo de transformação dos dados desta tabela é navegação e rotulagem dos dados, definindo atributos de data, texto ou número decimal. Este é resultante da integração com a pasta no *Sharepoint*, nele é utilizado informações do nome do arquivo também.
- BBCE: A transformação nessa tabela apresenta etapas seguindo a mesma lógica geral das anteriores, exceto pelo fato de possuir um atributo não normalizado, com diversas informações no mesmo atributo. Um exemplo pode ser visto na figura 42.

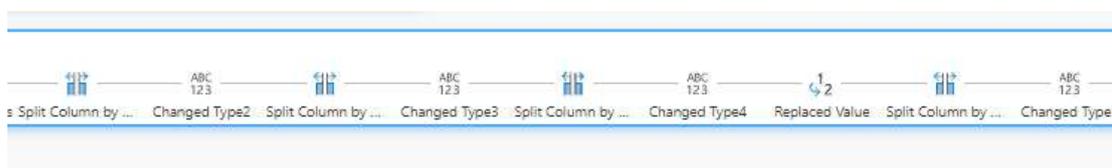
Figura 42 – Atributo Produto.

ABC 123	Produto
	SE CON ANU JAN/23 DEZ/23 - Preço Fixo
	SE CON ANU JAN/23 DEZ/23 - Preço Fixo
	SE CON ANU JAN/24 DEZ/24 - Preço Fixo
	SE CON ANU JAN/23 DEZ/23 - Preço Fixo
	SE CON ANU JAN/25 DEZ/25 - Preço Fixo
	SE CON ANU JAN/23 DEZ/23 - Preço Fixo
	SE CON ANU JAN/23 DEZ/23 - Preço Fixo
	SE CON ANU JAN/23 DEZ/23 - Preço Fixo
	SE CON ANU JAN/25 DEZ/25 - Preço Fixo
	SE CON ANU JAN/23 DEZ/23 - Preço Fixo
	SE CON ANU JAN/23 DEZ/23 - Preço Fixo
	SE CON ANU JAN/23 DEZ/23 - Preço Fixo
	SE CON ANU JAN/24 DEZ/24 - Preço Fixo

Fonte: Aluno.

Para esse processo foi utilizado a funcionalidade de dividir por delimitador, que transforma uma coluna em mais dependendo de um caractere selecionado. Resultando assim em Região, Tipo de energia, Tipo de Produto, Data de começo e fim da entrega e tipo de contrato. Etapas deste processo podem ser vistas na figura 43

Figura 43 – Etapas da transformação de dados da BBCE.



Fonte: Aluno.

- Preços PLD CCEE: Similar a etapas utilizadas nas demais tabelas, esse processo usa agrupamento com média para criar um preço com data de granularidade mensal não mais horária. A área de controle de risco não utiliza informações de granularidade horária de preços e sim mensal ou anual o que também diminui a quantidade de dados no conjunto.
- dCalendar: Além das tabelas já mapeadas foram criadas nessa ferramenta tabelas auxiliares, as quais ajudaram no desenvolvimento da fase de modelagem da metodologia. A tabela calendário é uma tabela padrão nos projetos no *Power BI*, ela gera datas diárias do ano 2018 até 2030 e disponibiliza diferentes formatos para data. Um esquema com as etapas desta transformação pode ser visualizado na figura 44

Figura 44 – Etapas de transformação dCalendar.



Fonte: Aluno.

- *Mapping*: Criação de tabelas para índices que ordenam dados dos visuais, tornam atributos comuns a outras tabelas através de relacionamentos.

#### DAX:

A Linguagem DAX é utilizada para criar medidas e colunas, pensadas para serem usadas no contexto do visual. Neste projeto ela é utilizada para criar indicadores de desempenho, medidas, conversão de unidade de medida e cálculos complexos. As medidas utilizadas são listadas abaixo:

- Dias Operados: Calcula a contagem de dias nos quais existem operações na tabela da BBCE, conforme mostra o código da figura 45.

Figura 45 – Código DAX: número de dias operados na BBCE.

```

Operated Days =
CALCULATE(
    COUNTAX(
        VALUES('data_MtM_BBCE'[Data Operação]),
        'data_MtM_BBCE'[Data Operação]
    ),
    (NOT(ISBLANK('data_MtM_BBCE'[Start Date])))
)

```

Fonte: Aluno.

Ele conta valor não brancos no atributo de Data da operação dependendo dos atributos utilizados no visual.

- Ano + 1: Essa medida calcula no contexto dos visuais, a soma de volume filtrada por tipo de produto anual e que o ano seja o ano do contexto mais um. Portanto, calcula a soma do volume do ano seguinte para dividir pelos dias operados trazendo assim um indicador de liquidez do mercado na BBCE. O código da medida é exposto na figura 46

Figura 46 – Código DAX: liquidez do produto do próximo ano.

```

1 Ano+1 =
2 VAR soma = CALCULATE(SUM('data_MtM_BBCE'[Volume]), 'data_MtM_BBCE'[Tipo de Produto] = "ANU"
3 && YEAR('data_MtM_BBCE'[Start Date]) = YEAR('data_MtM_BBCE'[Data Operação])+1
4 )
5 VAR conta =
6 CALCULATE(
7     COUNTAX(
8         VALUES('data_MtM_BBCE'[Data Operação]),
9         'data_MtM_BBCE'[Data Operação]
10    ),
11    (NOT(ISBLANK('data_MtM_BBCE'[Start Date])))
12 )
13 )
14
15 // VAR purchases = CALCULATE(SUM('data'[Volume]), 'data'[Activity] = "Purchases")
16 // VAR reserva = CALCULATE(SUM('data'[Volume]), 'data'[Way] = "Reserva")
17 RETURN
18 DIVIDE(soma, conta, 9999)

```

Fonte: Aluno.

- Nível de Contratação: Esse é um indicador utilizado para medir quanto de energia está contratado baseado na versão do Balanço. Ele calcula volume filtrado pelo

tipo de atividade, venda, compra ou geração, em forma de variáveis para realizar um cálculo conforme fórmula abaixo e figura do código 47:

$$1 - \frac{(\text{geração} + \text{compra} - \text{vendas})}{(\text{geração} + \text{compra})}$$

Figura 47 – Código DAX: Nível de contratação.

```
1 HCR =  
2 VAR sales = CALCULATE(SUM('datamatEBE'[Volume]), 'datamatEBE'[Activity] = "Sales")  
3 VAR generation = CALCULATE(SUM('datamatEBE'[Volume]), 'datamatEBE'[Activity] = "Generation")  
4 VAR purchases = CALCULATE(SUM('datamatEBE'[Volume]), 'datamatEBE'[Activity] = "Purchases")  
5 RETURN  
6 1 - DIVIDE(generation+purchases-sales, generation + purchases, 11)
```

Fonte: Aluno.

- Analogamente ao cálculo de nível de contratação a Posição Aberta é calculada pela soma das entradas de energia: geração e compras, menos saída de energia: vendas.

*Load:* Após a realização das outras duas ações do processo de ETL a etapa final é a de criação automática de um conjunto de dados em nuvem. Este será usado para alimentar o BI e será atualizado junto com o fluxo de dados.

#### 5.1.4 Modeling

Nesta etapa ainda no fluxo de dados são modelados os relacionamentos e agendados as atualizações automáticas. As atualizações foram agendadas semanalmente em dias úteis alternados conforme ilustrado na Figura 48.

Figura 48 – Configuração de agendamento da atualização.

Refresh

Configure a refresh schedule

Define a data refresh schedule to import data from the data source into the semantic model. [Learn more](#)

On

Refresh frequency

Weekly

Time zone

(UTC-03:00) Brasilia

Sunday

Monday

Tuesday

Wednesday

Thursday

Friday

Saturday

Time

[Add another time](#)

Send refresh failure notifications to

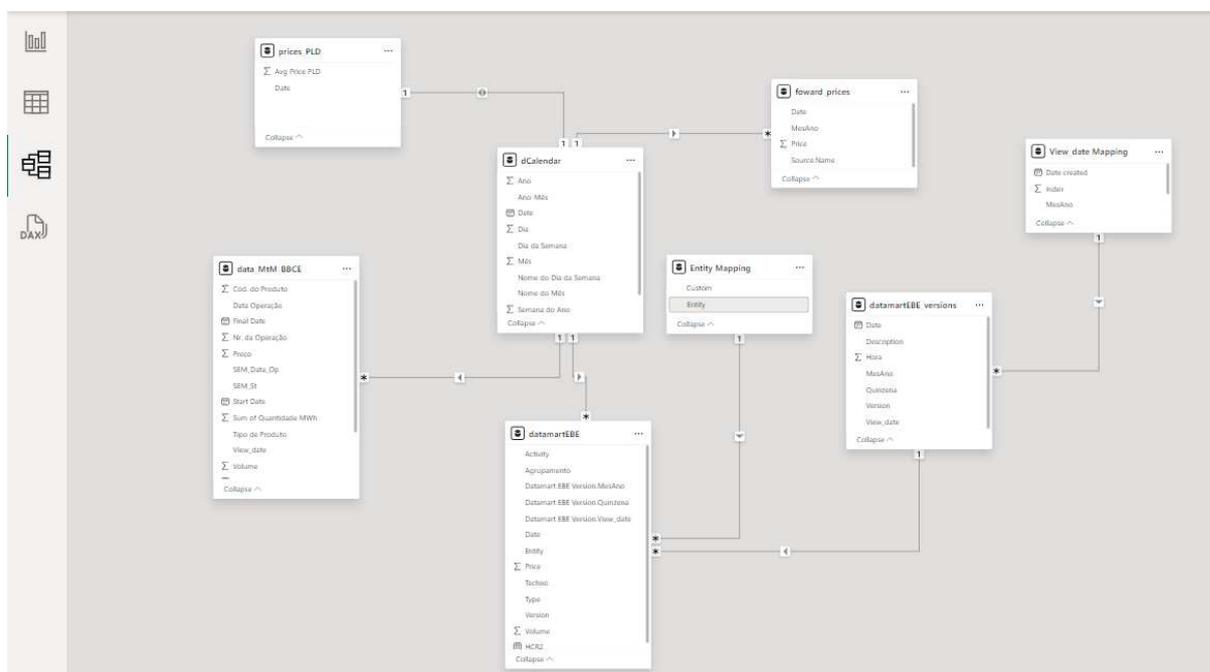
Dataflow owner

These contacts:

Fonte: Aluno.

Os relacionamentos foram modelados de modo que atributos de diferentes tabelas podem se comunicar, essa etapa deve fazer sentido com os objetivos de navegabilidade, entendimento e interação do cliente final ao *Power BI*. Os relacionamentos do modelo de dados são apresentados na figura 49.

Figura 49 – Relacionamentos Data Mart Controle de Risco.



Fonte: Aluno.

Na figura fica evidente os relacionamentos de mapeamento para ordenar visuais, de data comum a mais tabelas e outros formatos de datas. Os relacionamentos utilizados nessa modelagem são de um pra muitos, onde existe apenas um registro do atributo em uma das tabelas.

Páginas do relatório:

- **Portfólio:** O objetivo desta página do relatório é demonstrar um panorama geral do portfólio do grupo Engie, ele é demonstrado através de um gráfico de colunas empilhadas que cada categoria do eixo x mostra a divisão por atividades de compra, venda e geração. Também cada coluna empilhada é subdividida pelos valores do atributo de tecnologia da energia. Esta página é apresentada na figura 50.

Figura 50 – Página Portfólio do relatório de Controle de Risco.



Fonte: Aluno.

Para representar a medida DAX da posição aberta da empresa, é utilizado um visual de cartão e uma tabela para detalhar os valores utilizados para o cálculo dela. Um visual de venda por tipo de mercado também é utilizado para dar uma visão geral do tipo de vendas da energia do grupo. Todos esses visuais são interativos podendo focar em um valor refletindo nos outros visuais, navegar alterando visuais de filtros de acordo com a versão e ano desejada.

- **Liquidez BBCE** Os visuais desta página apresentam uma visão sobre a liquidez do mercado livre na plataforma da BBCE. Essa visão é importante para compreender qual e quanto dos produtos estão sendo negociados, apoiando assim em decisões estratégicas que impactam no lucro do grupo.

Figura 51 – Página Liquidez na BBCE do relatório de Controle de Risco.



Fonte: Aluno.

A página é composta por uma tabela que mostra de acordo com o mês e ano da data de operação, quantos dias existiram operações. Também conta com um filtro de ano para navegar entre as informações dos visuais e um gráfico com data de operação mensal no eixo horizontal mostrando medidas. Estas medidas demonstram através de colunas a soma de volume do produto com entrega no ano ou mês, seguinte ou mais no futuro dependendo dos parâmetros escolhidos. A página de liquidez da BBCE é apresentada na figura 51

- **Nível de Contratação** A página indica o Nível de Contratação através de um gráfico de linhas, construído com valores mensais da versão da tabela Balanço Unificado com mês e ano da versão no eixo horizontal e a medida de nível de contratação em porcentagem como uma das linhas do gráfico. A outra linha é o nível de contratação planejado da empresa, sendo um cenário ruim estar abaixo desta linha. Nesse gráfico é adicionado um filtro do ano atual e mais dois anos pra frente automático que fica dentro do visual, não permitindo interação nesse gráfico já que ele sempre é baseado no momento atual. Essa página é apresentada na figura 52.

Figura 52 – Página Nível de Contratação do relatório de Controle de Risco.



Fonte: Aluno.

- Preços O objetivo desta página é trazer uma visão histórica, conforme o usuário interage alterando os filtros, do preço tem uma visão de cenários de preços futuros conforme a expectativa da empresa e do preço de liquidação na CCEE (PLD), ou seja, o preço de venda de energia no mercado regulado. Esta página é apresentada na figura 53

Figura 53 – Página Preços do relatório de Controle de Risco.



Fonte: Aluno.

O gráfico de linha do PLD usa datas mensais no eixo horizontal e a média do preço PLD como linha tendo um filtro do último ano. O outro gráfico de linha mostra o preço mensal dos meses e anos futuro de acordo com a versão selecionada no filtro e de uma visão do orçamento chamada "V1/23".

### 5.1.5 Avaliação e Implantação

Nesta etapa foi realizada uma avaliação prévia do resultado da solução pela equipe de desenvolvimento. Nesta avaliação, feita pela equipe de desenvolvimento, conclui-se que o projeto está adequado em termos de eficiência, com bom desempenho e correção de dados, além da usabilidade do funcionamento do relatório. Consolidado o sistema do *Power BI*, publicou-se o projeto e liberou-se o acesso a alguns gestores para o sistema ser testado por usuário sem conhecimento do desenvolvimento ao mesmo tempo que foram coletadas opiniões críticas ao sistema. Nesta etapa é realizado a classificação de acesso garantindo segurança dos dados sigilosos de acordo com cada credencial.

No capítulo seguinte, a análise de resultados será detalhada.

## 6 ANÁLISE DE RESULTADOS

Este capítulo realiza uma análise crítica e profunda dos resultados obtidos através da implantação do projeto, demonstrando as atividades realizadas, os resultados alcançados, os impactos na organização, vantagens e desvantagens. Também é elucidado como a solução resolveu os desafios propostos avaliando resultados através da validação e testes. Ele é dividido em duas partes, sendo a arquitetura de dados a primeira e a inteligência de negócio através de relatórios na ferramenta Power BI a segunda.

### 6.1 RESULTADO ARQUITETURA DE DADOS

A parte da arquitetura de dados é vital para o projeto, ela é responsável pelo fluxo de dados que serve de base para a segunda parte. A qualidade desta parte é diretamente proporcional à qualidade dos relatórios de BI. Não só a qualidade como também ao funcionamento desta a longo do ciclo de vida do projeto. A simplificação da arquitetura, centralizando o processo de ETL no ambiente online, que integra as fontes de dados cortando passos intermediários, torna o processo mais fácil de se entender, menos vulnerável a falhas como arquivos corrompidos e mais assertivo pulando etapas intermediárias desnecessárias em ambientes diferentes. Na parte de arquitetura de dados também pode ser mensurado a diminuição do tempo de trabalho da equipe que era de um dia, quando não existia necessidade de apagar os dados manualmente para reprocessar por algum erro, para apenas o tempo de adicionar o arquivo à pasta compartilhada de menos de dois minutos duas vezes por mês.

- Fontes de dados: A replicação do repositório de dados de outro domínio: Dados extraídos do domínio de outra entidade do grupo, servindo de fonte de dados na arquitetura. A criação de uma pasta como fonte de dados não é a maneira mais indicada para servir de banco de dados, porém ela é necessária pelo fato dessas informações não estarem organizadas e disponíveis para integração. Também é importante para manter o projeto acessível a todos os membros da equipe, sendo esta solução mais amigável a leigos em tecnologia. A replicação do *Data Warehouse* poupa o sistema de reprocessar um grande volume de dados, o que não impacta positivamente o desempenho da arquitetura, além de dar governança e ser uma ponte de acesso para novos conjuntos de dados. Ao conectar as fontes pulando etapas intermediárias em outros ambientes, tem-se a melhora na assertividade dos dados comparado com os dados gerais utilizados pela empresa. Outra vantagem é a melhora no desempenho diminuindo integrações com ambientes diferentes e o tempo de trabalho da equipe que agora não precisa construir dados manualmente em diversas fontes de ambientes distintos;

- **Processo de ETL:** O processo de ETL ocorre no fluxo de dados, todas as fontes de dados estão integradas a arquitetura de maneira segura e consistente através *gateways* que são responsáveis pela liberação dessa conexão. Os dados são transformados pela ferramenta *online* possibilitando mais colaboradores simultâneos no processo como um todo. Estes dados são limpos, rotulados, filtrados e manipulados de maneira que fiquem adequados para análises sem erros e sem dados não utilizados. Assim torna o processo de ETL e de carregamento na ferramenta mais leves e conseqüentemente mais rápidos.

Carregando os dados para criação do conjunto foram realizados testes medindo o tempo que o fluxo com dados reais toma para atualizar. O tempo de atualização foi adequado, sendo ele menor que 60 segundos de processamento. Esse tempo tende a aumentar conforme adição de informações o que pode ser facilmente resolvido adicionando, na ferramenta de ETL, filtros com parâmetros de datas mais recentes. Comparando o conjunto de dados antigo contra o novo utilizando a nova arquitetura, notou-se pequenas diferenças, que na verdade eram inconsistências nos valores de volume calculados de outra maneira com processos descentralizados. Quando comparado aos dados da empresa que agora também são fontes da arquitetura os erros foram extintos.

O processamento de ETL centralizado online melhora o desempenho no processamento de informações, carregando o conjunto de dados já previamente processado. Este processo agora tem uma única versão da verdade e pode-se trabalhar no sistema de forma conjunta, diminuindo vulnerabilidade a falhas de processos de copiar e colar dados nas tabelas, corromper algum arquivo ou modificação de parâmetros manuais incorretamente. Além disso, ele é facilmente escalável integrando novas fontes de dados e caso ocorra algum erro é possível checar o sistema em um único ambiente;

- **Modelagem dados relacionais:** Após a transformação dos dados é realizada a modelagem das tabelas através de relacionamentos dependendo do objetivo, esses relacionamentos se incorretamente modelados são refletidos em falhas nas construções dos relatórios, o que não acontece.

A modelagem se realizada corretamente, melhora o desempenho da arquitetura diminuindo a redundância de dados o que torna o volume de dados menor e também impacta positivamente a usabilidade com a relação entre atributos de tabelas diferentes;

- **Atualização:** A taxa de atualização dos dados é agendada para acontecer três dias por semana, a atualização do relatório e do modelo de dados é sincronizada a ela. A atualização antes além de, necessitar intervenção humana uma vez por mês, tomava muito tempo para executar todo o processo de ETL em diversos

arquivos. Agora ela é feita de forma automática e combina com estratégias no ETL acrescenta duas versões por mês, disponibilizando dados mais recentes o que gera valor ao objetivo do projeto.

Esse processo de atualização que durava aproximadamente um dia de trabalho da equipe agora é realizada no ambiente online de forma agendada. Também é seguro pela classificação de acesso de colaboradores ou times de acordo com as necessidades ;

## 6.2 RESULTADOS RELATÓRIOS *POWER BI*

A parte final do projeto é de implantação do relatórios de *Business Intelligence*, ela é o final do fluxo de dados da arquitetura desenvolvida neste projeto. Foram reconstruídos os gráficos de acordo o padrão da empresa e adicionando visuais de filtros para navegação. Com objetivo de avaliar o relatório, foi disponibilizada a versão de avaliação para os integrantes da equipe de controle de risco, diretores e outros colaboradores que utilizam o relatório. Os *feedbacks* foram coletados por meio de um formulário baseado na escala Likert, onde os indicadores foram avaliados com valores de um a cinco, sendo um o mínimo grau de satisfação e cinco o máximo grau. Os tópicos do formulário incluíram usabilidade e performance, subdividindo avaliações específicas para cada página. O conteúdo do resultado do formulário é disponibilizado na tabela 54:

Figura 54 – Tabela resultado formulário de feedback.

Usabilidade	Cargo	Nota Pg-1	Nota Pg-2	Nota Pg-3	Nota Pg-4
	Gerente do Time	5	5	5	4.5
	Chefe TI	5	5	5	4.5
	Colaborador TI	5	5	5	5
	Analista do Time	4.5	5	5	5
	Coordenador do Time	5	5	5	5
	Usuário final outro time	5	5	5	5

Performance	Cargo	Nota Pg-1	Nota Pg-2	Nota Pg-3	Nota Pg-4
	Gerente do Time	5	5	5	5
	Chefe TI	4.5	4.5	5	4
	Colaborador TI	5	5	5	4.5
	Analista do Time	5	5	5	5
	Coordenador do Time	5	5	5	5
	Usuário final outro time	5	5	5	4.5

Fonte: Aluno.

O resultado do formulário reflete o alto grau de satisfação em relação as páginas

do relatório, proporcionando uma visão clara sobre tópicos com bom resultado e aspectos que podem indicar a necessidade de revisão e melhorias. Com média de todos os resultados sendo 4.91, indica que o resultado geral de projeto foi aprovado pelos usuários alvo, entretanto ele sempre estará disponível para mudanças e melhorias contínuas.

## 7 CONCLUSÃO

Neste capítulo final, serão apresentados e discutidos principais pontos do projeto. Ele será composto uma visão abrangente do projeto, discutindo os resultados globais, as limitações da solução desenvolvida, os pressupostos assumidos e sugestões para trabalhos futuros que possam dar continuidade e aprimorar os resultados alcançados.

O projeto foi teve como objetivo principal a reestruturação de uma arquitetura de dados do processo de *Business Intelligence* (BI) na área financeira de controle de risco da *holding* do grupo ENGIE no Brasil. Foram mapeadas diversas fontes de dados que compõem este processo de BI, com objetivo de integra-las em um conjunto de dados *online* e centralizado através do processo de ETL. No mapeamento, descobriu-se que a maioria das fontes de dados utilizadas eram de domínio de outra entidade do grupo.

A partir do fato descoberto, foram selecionados os conjuntos de dados do domínio comuns as fontes utilizadas pela arquitetura e replicado o *Data Warehouse* (DW) contendo os conjuntos desejados. A replicação deste DW permitiu aos administradores, controle total sobre possíveis mudanças na estrutura dos conjuntos de dados, aumentando assim a segurança a falhas e acurácia dos dados.

As fontes de dados restantes foram modeladas como uma pasta em nuvem no Sharepoint, esta estratégia garante acessibilidade aos membros do time não especializados na área de tecnologia e cria um modelo fácil para inserção de análises pontuais sem precisar se envolver em todo o processo de ETL.

Utilizando a ferramenta de ETL Power Query Editor *online* através do *Data Flow*, os dados foram extraídos, transformados e carregados em um conjunto de dados disponível *online*. Esta solução aumenta a colaboratividade, já que disponibiliza mais de um acesso as ferramentas utilizadas no projeto. Também aumenta a performance do processo de BI como um todo, já que os dados são processados na nuvem e sem outros processos de ETL intermediários e descentralizados. Esses processos agora encontram-se centralizados na ferramenta aumentando o entendimento do fluxo de dados, garantindo eficiência do processo e aumentando a precisão dos dados nele contidos.

O uso do *Data Flow* possibilita automatizar o processo de ETL aumentando a qualidade das informações através de informações mais recentes e relevantes. Nele também são criados relacionamentos das tabelas aumentando poder de análise do conjunto de dados que serão utilizados na criação dos relatórios interativos contendo conhecimento estratégico do negócio.

Os resultados globais do projeto foram muito satisfatórios. Na parte da arquitetura de dados, a qualidade e a eficiência no processamento foram significativamente aprimoradas. A integração e processamento dos dados de diversas fontes eliminaram

pequenas inconsistências garantindo a assertividade das informações. A automatização dos processos de ETL reduziu significativamente o tempo de processamento na atualização dos dados além de possibilitar o agendamento dessa função.

Na outra parte dos relatórios de *Business Intelligence*, as páginas contendo visuais interativos disponíveis a qualquer momento através da ferramenta Power BI, melhoraram a experiência do usuário, facilitando a navegação e a obtenção de conhecimento. O *feedback* coletado de usuários finais, colegas da área de controle de rico e da área de tecnologia da informação, demonstrou um alto grau de satisfação, com uma média de 4.91 em uma escala de 1 a 5, o que valida a satisfação e eficácia da nova solução desenvolvida.

Apesar dos resultados positivos, o projeto apresentou algumas limitações. A integração de dados através de uma pasta no Sharepoint, embora acessível, pode ser vulnerável a erros e requer interação humana. A integração de múltiplas fontes de dados apresentou desafios, especialmente no mapeamento das diferentes fontes de dados e dos etapas de transformação dos dados para manter a consistência.

Durante o desenvolvimento do projeto, alguns pressupostos foram assumidos. Assumiu-se que os recursos necessários, como ferramentas de software e infraestrutura de TI, estariam disponíveis durante todo o processo e fossem eficientes e compatível com a nova arquitetura de dados. Também foi verificada a segurança do sistema através do gerenciamento de acessos de cada etapa do processo garantindo que só colaboradores com credenciais registradas tenham acesso para proteger a integridade e a confidencialidade dos dados sensíveis.

Para dar continuidade e aprimorar os resultados alcançados, é sugerido algumas melhorias futuras para o projeto. A automação completa dos processos de atualização de dados pode eliminar a necessidade de intervenções manuais, aumentando ainda mais a eficiência e diminuindo tempo de trabalho dos colaboradores. A integração de novas fontes de dados pode expandir a abrangência e a riqueza das análises de BI. Além disso, a implementação de um sistema de monitoramento contínuo pode avaliar o desempenho da nova arquitetura alertando a necessidade de manutenção conforme necessário.

Por fim, o projeto mostrou-se desafiador, entretanto ele alcançou seu objetivo principal de melhorar a arquitetura de dados e o sistema de BI, trazendo benefícios significativos em termos de eficiência, qualidade dos dados e satisfação dos usuários. Além das limitações identificadas, que servem de base para melhorias futuras, é importante desenvolver um sistema de *feedback* contínuo para garantir que a solução continue a evoluir e atender às necessidades de inteligência de negócio da empresa. A continuidade dos trabalhos futuros e a adoção das sugestões propostas irão assegurar que a empresa mantenha uma posição relevante e competitiva no competitivo setor energético.

## REFERÊNCIAS

ABRADEE. **Visão Geral do Setor**. [S.l.: s.n.], 2024. Acessado em: 7 de junho de 2024. Disponível em: <https://abradee.org.br/visao-geral-do-setor/>.

AMAZON. **O que é a computação em nuvem?** Disponível em: <https://aws.amazon.com/pt/what-is-cloud-computing/>.

AMAZON WEB SERVICES. **O que é um banco de dados?** [S.l.: s.n.], 2024. Acesso em: 7 jun. 2024. Disponível em: <https://aws.amazon.com/pt/what-is/database/>.

ANDREATTO, Ricardo. **Construindo um Data Warehouse e analisando suas informações com Data Mining e OLAP**. 1999. Monografia de Conclusão – Valinhos.

BALCÃO BRASILEIRO DE COMERCIALIZAÇÃO DE ENERGIA. **Plataforma Balcão Eletrônico**. Acessado em 7 de junho de 2024. s.d. Disponível em: <https://www.bbce.com.br/plataforma-balcao/#:~:text=A%20plataforma%20Balc%C3%A3o%20Eletr%C3%B4nico%20BBCE,e%20a%20volatilidade%20de%20pre%C3%A7os.>

BARBOSA FILHO, Luiz Henrique. **Design de Banco de Dados**. Publicado em: 27/09/2022. Análise Macro. 2022. Disponível em: <https://analisemacro.com.br/data-science/design-de-banco-de-dados/>.

BONFIOLI, Guilherme Ferreira. **Banco de Dados Relacional e Objeto-Relacional: uma comparação usando PostgreSQL**. 2006. Monografia de graduação – Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil.

BORGES, Hélder Pereira; SOUZA, José Neuman de; SCHULZE, Bruno; MURY, Antonio Roberto. **Computação em Nuvem**. [S.l.: s.n.], 2011.

BRUNI, Adriano Leal; GOMES, Sônia Maria da Silva (Ed.). **Controladoria Empresarial: Conceitos, ferramentas e desafios**. 1. ed. Salvador: Edufba, 2010. P. 44.

CÂMARA DE COMERCIALIZAÇÃO DE ENERGIA ELÉTRICA. **Painel de Preços**. [S.l.: s.n.]. Acesso em: 10 jun. 2024. Disponível em: <https://www.ccee.org.br/web/guest/precos/painel-precos>.

- CASTRO, Nivalde José de; OSÓRIO, Mauro; REGO, André. **Avaliação da Viabilidade Econômica da Implementação de Veículos Elétricos no Brasil**. [S.l.], 2021. Acessado em: 7 de junho de 2024. Disponível em: [https://gesel.ie.ufrj.br/wp-content/uploads/2022/06/03\\_castro\\_2021\\_03\\_08.pdf](https://gesel.ie.ufrj.br/wp-content/uploads/2022/06/03_castro_2021_03_08.pdf).
- CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. **CRISP-DM 1.0: Step-by-step data mining guide**. [S.l.], 2000.
- CODD, E. F. A Relational Model of Data for Large Shared Data Banks. **Communications of the ACM**, IBM Research Laboratory, San Jose, California, v. 13, n. 6, p. 377–387, 1970.
- CORPORATE FINANCE INSTITUTE. **Data Warehousing**. Disponível em: <https://corporatefinanceinstitute.com/resources/business-intelligence/data-warehousing/>.
- DOMINGUES, R.; PEDROSA, I.; BERNARDINO, J. **Indicadores Chave de Desempenho em Marketing**, 2021.
- ENERGIA ELÉTRICA (ANEEL), Agência Nacional de. **A ANEEL**. [S.l.: s.n.], 2024. Acessado em: 7 de junho de 2024. Disponível em: <https://www.gov.br/aneel/pt-br/acesso-a-informacao/institucional/a-aneel>.
- IBM. **Bancos de Dados Relacionais**. [S.l.: s.n.], 2024a. Acesso em: 7 jun. 2024. Disponível em: <https://www.ibm.com/br-pt/topics/relational-databases>.
- IBM. **Estrutura de Relacionamento de Banco de Dados**. Disponível em: <https://www.ibm.com/docs/pt-br/control-desk/7.6.1.2?topic=structure-database-relationships>.
- IBM. **Visão geral da ajuda do CRISP-DM**. [S.l.: s.n.], 2024b. Acessado em: 10 de maio de 2024. Disponível em: <https://www.ibm.com/docs/pt-br/spss-modeler/18.5.0?topic=dm-crisp-help-overview>.
- JESUS, Andressa Costa de. **Business Intelligence: Um Estudo de Sua Aplicação Como Ferramenta de Apoio a Micro e Pequenas Empresas**. São Mateus: Instituto Vale do Cricaré, Faculdade Vale do Cricaré, Curso de Análise e Desenvolvimento de Sistemas, 2020.

JESUS, Andressa Costa de. **Business Intelligence: Um Estudo de Sua Aplicação Como Ferramenta de Apoio a Micro e Pequenas Empresas**. 2020. Instituto Vale do Cricaré, Faculdade Vale do Cricaré, São Mateus.

KAPLAN, Robert S.; NORTON, David P. **A Estratégia em Ação: Balanced Scorecard**. Tradução: Luiz Euclides Trindade Frazão Filho. 5. ed. Rio de Janeiro: Campus, 1997.

KIMBALL, Ralph. **Data Warehouse Toolkit - Guia Completo Para Modelagem Dimensional**. [S.l.]: Editora Campus, 2002.

KIYAN, F.M. **Indicadores de Desempenho como Suporte Estratégico**. 2001. F. 118. Dissertação de Mestrado – Universidade de São Paulo, São Paulo.

LORENZETTI, C. **Business Intelligence Systems in the Financial Industry**. Italy: [s.n.], 2010. Citado na página 16.

MICROSOFT. **Cenário de exemplo: Modelagem de desempenho e confiabilidade de site para aplicativos escaláveis**. Disponível em:

<https://learn.microsoft.com/pt-br/azure/architecture/example-scenario/apps/scalable-apps-performance-modeling-site-reliability>.

MICROSOFT. **Extract, Transform, and Load (ETL)**. [S.l.: s.n.], 2024. Acesso em: 7 jun. 2024. Disponível em: <https://learn.microsoft.com/pt-br/azure/architecture/data-guide/relational-data/etl>.

MICROSOFT. **Gateway de Serviço local para integração de dados**. Disponível em: <https://learn.microsoft.com/pt-br/data-integration/gateway/service-gateway-onprem>.

MICROSOFT. **O que é Power Query?** Disponível em: <https://learn.microsoft.com/pt-br/power-query/power-query-what-is-power-query>.

MICROSOFT. **Visão geral do Datamarts**. Disponível em: <https://learn.microsoft.com/pt-br/power-bi/transform-model/datamarts/datamarts-overview>.

MICROSOFT. **Visão geral do Power BI**. Disponível em: <https://learn.microsoft.com/pt-br/power-bi/fundamentals/power-bi-overview>.

MICROSOFT AZURE. **Infraestrutura Global do Microsoft Azure**. Disponível em: <https://azure.microsoft.com/pt-br/explore/global-infrastructure/>.

MICROSOFT AZURE. **O que são bancos de dados?** [S.l.: s.n.], 2024. Acesso em: 7 jun. 2024. Disponível em: <https://azure.microsoft.com/pt-br/resources/cloud-computing-dictionary/what-are-databases>.

NAVATHE, Shamkant B.; ELMASRI, Ramez E. **Sistemas de Banco de Dados**. 4. ed. [S.l.]: Editora, se aplicável, 2005.

O'BRIEN, James A. **Sistemas de informação e as decisões gerenciais na era da internet**. 3. ed. São Paulo: Saraiva, 2010.

OLIVEIRA, L. M.; PEREZ JÚNIOR, J. H.; SILVA, C. A. S. **Controladoria Estratégica**. 5. ed. São Paulo: Atlas, 2009.

OPERADOR NACIONAL DO SISTEMA ELÉTRICO (ONS). **O que é o Sistema Interligado Nacional (SIN)?** Acessado em 7 de junho de 2024. 2024. Disponível em: <https://www.ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin>.

ORACLE. **O que é um Banco de Dados?** [S.l.: s.n.], 2024. Acesso em: 7 jun. 2024. Disponível em: <https://www.oracle.com/br/database/what-is-database/>.

PEREIRA, Cassia Gabriela Silva. **Análise de Dados na Saúde Pública: Um Estudo de Caso da Hanseníase no Estado do Tocantins**. 2021. Trabalho de Conclusão de Curso – Universidade Federal do Tocantins, Palmas, Tocantins.

PESQUISA ENERGÉTICA (EPE), Empresa de. **Matriz Energética e Elétrica**. [S.l.: s.n.], 2024. Acessado em: 7 de junho de 2024. Disponível em: <https://www.epe.gov.br/pt/abcdenergia/matriz-energetica-e-eletrica>.

PIATETSKY, Gregory. Crisp-dm, still the top methodology for analytics, data mining, or data science projects. **KDD News**, 2014. Citado na página 10.

PRIMAK, Fábio Vinícius. **Decisões com BI (Business Intelligence)**. [S.l.]: Editora Ciência Moderna, 2008. v. 1.

PRIMAK, Fabio Vinícius. **Decisões com BI (Business Intelligence)**. [S.l.]: Editora Ciência Moderna, 2008. v. 1, p. 22.

RAMAKRISHNAN, Raghu; GEHRKE, Johannes. **Sistema de Gerenciamento de Banco de Dados**. 3. ed. New York: McGraw-Hill, 2011.

SERIAN, João Sidemar. **Business Intelligence: Cresce o mercado para profissionais de BI**. São Paulo: [s.n.], 2007. Acesso em: 7 jun. 2024. Disponível em: [http://www.imasters.com.br/artigo/5354/bi/cresce\\_o\\_mercado\\_para\\_profissionais\\_de\\_bi/](http://www.imasters.com.br/artigo/5354/bi/cresce_o_mercado_para_profissionais_de_bi/).

SIQUEIRA, Tagore Villarim de. O setor de tecnologia da informação e comunicação no Brasil no período recente. **Revista do BNDES**, v. 14, n. 27, p. 213–260, 2007.

SISTEMA ELÉTRICO (ONS), Operador Nacional do. **O que é ONS**. [S.l.: s.n.], 2024. Acessado em: 7 de junho de 2024. Disponível em: <https://www.ons.org.br/paginas/sobre-o-ons/o-que-e-ons>.

TEIXEIRA, Maíra. O futuro digital das finanças. **Deloitte**, v. 56, 2017.

TURBAN, Efraim; SHARDA, Ramesh; ARONSON, Jay E.; KING, David. **Business Intelligence - Um enfoque Gerencial para a inteligência do Negócio**. Porto Alegre: Bookman, 2009.

VILELA, Tarek Mohamad. **Precificação de opções no mercado de energia elétrica brasileiro**. [S.l.: s.n.], 2022. Monografia (Graduação em Economia). Fundação Getulio Vargas, Escola de Economia de São Paulo, São Paulo.

WEISS, Andressa Thais. **A Controladoria como Ferramenta de Gestão para Micro e Pequenas Empresas Comerciais do Ramo de Vestuário**. 2013. Trabalho de Conclusão de Curso – Universidade Federal de Santa Maria, Santa Maria, RS, Brasil. Orientador: Prof. Msc. Gilberto Brondani.

ZDANOWICZ, José Eduardo. **Fluxo de Caixa: Uma Decisão de Planejamento e Controle Financeiro**. 3. ed. Porto Alegre: D. C. Luzzatto, 1989.

# **Apêndices**

**APÊNDICE A – CÓDIGOS M DAS ETAPAS DE CRIAÇÃO DAS TABELAS DO  
CONJUNTOS DE DADOS**

```
//Etapas Obtenção tabela BBCE #parte 1
let
    // Conexão e seleção da tabela
    Source = AmazonRedshift.Database(
        "your-cluster-endpoint.amazonaws.com",
        "your-database-name",
        [
            CredentialConnectionString = [
                UserName = "your-username",
                Password = "your-password",
                Encrypted = false,
                Host = "your-cluster-endpoint.amazonaws.com",
                Database = "your-database-name",
                Port = "5439"
            ]
        ]
    ),
    Schema = Source([Name="your-schema-name",Kind="Schema"])[Data],
    Table = Schema([Name="your-table-name",Kind="Table"])[Data],

    //Fim Genérico Extração e navegação
    #"Promoted Headers" = Table.PromoteHeaders(Table, [PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"Date", type
date}}),
    #"Replaced Errors" = Table.ReplaceErrorValues(#"Changed Type", {"Date", #date(2022, 3,
1)}),
    #"Replaced Errors1" = Table.ReplaceErrorValues(#"Replaced Errors", {"Date", #date(2022, 3,
1)}),
    Filtro_Dias = Table.SelectRows(#"Replaced Errors1", each
    let
        dia = Date.Day([Date]),
        desc = [Description]
```

```
/Etapas Obtenção tabela BBCE #parte 2
in
    ((dia >= 3 and dia <= 5) or (dia >= 13 and dia <= 15)) or desc = "budget"

),
Select_newest = Table.AddColumn(Filtro_Dias, "Quinzena", each
    let
        dia = Date.Day([Date])
    in
        if dia >= 3 and dia <= 5 then 1
        else if dia >= 13 and dia <= 15 then 2
        else null),
MesAno = Table.AddColumn(Select_newest, "MesAno", each
    let
        mes = Text.Start(Date.ToText([Date], "MMM"), 3),
        ano = Text.End(Text.From(Date.Year([Date])), 2)
    in
        mes & "/" & ano, type text),
#"Grouped Rows" = Table.Group(MesAno, {"Date"}, {"Count", each Table.RowCount(_),
Int64.Type}, {"Todas Linhas", each _ type table [Version=number, Date=nullable date,
Hora=number, Description=text, Quinzena=number, MesAno=text]}, {"Max hora", each
List.Max([Hora]), type number})),
#"Expanded Todas Linhas" = Table.ExpandTableColumn("#Grouped Rows", "Todas Linhas",
{"Version", "Hora", "Description", "Quinzena", "MesAno"}, {"Version", "Hora", "Description",
"Quinzena", "MesAno"}),
#"Filtro Max hora" = Table.SelectRows("#Expanded Todas Linhas", each [Hora] = [Max hora] ),
#"Removed Other Columns" = Table.SelectColumns("#Filtro Max hora", {"Date", "Version",
"Hora", "Description", "Quinzena", "MesAno"}),
#"Grouped Rows1" = Table.Group("#Removed Other Columns", {"Quinzena", "MesAno"},
{"Count", each Table.RowCount(_), Int64.Type}, {"All Rows", each _ type table [Date=nullable
date, Version=nullable number, Hora=nullable number, Description=nullable text,
Quinzena=nullable number, MesAno=nullable text]}, {"Max Date by Day and Quinzena", each
List.Max([Date]), type nullable date})),
#"Expanded All Rows" = Table.ExpandTableColumn("#Grouped Rows1", "All Rows", {"Date",
"Version", "Hora", "Description"}, {"Date", "Version", "Hora", "Description"}),
```

```
/Etapas Obtenção tabela BBCE #parte 3

#"Filtro Max Day_Quinzena" = Table.SelectRows("#Expanded All Rows", each [Date] = [Max
Date by Day and Quinzena] ),

#"Removed Other Columns1" = Table.SelectColumns("#Filtro Max
Day_Quinzena",{"Quinzena", "MesAno", "Date", "Version", "Hora", "Description"}),

#"Changed Type1" = Table.TransformColumnTypes("#Removed Other
Columns1",{{"Quinzena", type text}}),

#"Added Custom" = Table.AddColumn("#Changed Type1", "View_date", each [MesAno] &"-
"&[Quinzena])

in

#"Added Custom"
```

```
// Etapas Balanço Unificado Versões #parte 1

let

// Conexão e seleção da tabela

Source = AmazonRedshift.Database(

"your-cluster-endpoint.amazonaws.com",

"your-database-name",

[

CredentialConnectionString = [

UserName = "your-username",

Password = "your-password",

Encrypted = false,

Host = "your-cluster-endpoint.amazonaws.com",

Database = "your-database-name",

Port = "5439"

]

]

),

Schema = Source[{Name="your-schema-name",Kind="Schema"}][Data],

Table = Schema[{Name="your-table-name",Kind="Table"}][Data],
```

```
// Etapas Balanço Unificado Versões #parte 2
//Fim Genérico Extração e navegação
#"Promoted Headers" = Table.PromoteHeaders(Table, [PromoteAllScalars=true]),
#"Changed Type" = Table.TransformColumnTypes("#Promoted Headers",{{"Date", type
date}}),
#"Replaced Errors" = Table.ReplaceErrorValues("#Changed Type", {"Date", #date(2022, 3,
1)}),
#"Replaced Errors1" = Table.ReplaceErrorValues("#Replaced Errors", {"Date", #date(2022, 3,
1)}),
Filtro_Dias = Table.SelectRows("#Replaced Errors1", each
let
    dia = Date.Day([Date]),
    desc = [Description]
in
    ((dia >= 3 and dia <= 5) or (dia >= 13 and dia <= 15)) or desc = "budget"
),
Select_newest = Table.AddColumn(Filtro_Dias, "Quinzena", each
let
    dia = Date.Day([Date])
in
    if dia >= 3 and dia <= 5 then 1
    else if dia >= 13 and dia <= 15 then 2
    else null),
MesAno = Table.AddColumn(Select_newest, "MesAno", each
let
    mes = Text.Start(Date.ToText([Date], "MMM"), 3),
    ano = Text.End(Text.From(Date.Year([Date])), 2)
in
    mes & "/" & ano, type text),
#"Grouped Rows" = Table.Group(MesAno, {"Date"}, {"Count", each Table.RowCount(_),
Int64.Type}, {"Todas Linhas", each _, type table [Version=number, Date=nullable date,
```

```

// Etapas Balanço Unificado Versões #parte 3

Hora=number, Description=text, Quinzena=number, MesAno=text], {"Max hora", each
List.Max([Hora]), type number}}),

    #"Expanded Todas Linhas" = Table.ExpandTableColumn("#Grouped Rows", "Todas Linhas",
{"Version", "Hora", "Description", "Quinzena", "MesAno"}, {"Version", "Hora", "Description",
"Quinzena", "MesAno"}),

    #"Filtro Max hora" = Table.SelectRows("#Expanded Todas Linhas", each [Hora] = [Max hora] ),

    #"Removed Other Columns" = Table.SelectColumns("#Filtro Max hora",{"Date", "Version",
"Hora", "Description", "Quinzena", "MesAno"}),

    #"Grouped Rows1" = Table.Group("#Removed Other Columns", {"Quinzena", "MesAno"},
{{"Count", each Table.RowCount(_), Int64.Type}, {"All Rows", each _, type table [Date=nullable
date, Version=nullable number, Hora=nullable number, Description=nullable text,
Quinzena=nullable number, MesAno=nullable text]}, {"Max Date by Day and Quinzena", each
List.Max([Date]), type nullable date}}),

    #"Expanded All Rows" = Table.ExpandTableColumn("#Grouped Rows1", "All Rows", {"Date",
"Version", "Hora", "Description"}, {"Date", "Version", "Hora", "Description"}),

    #"Filtro Max Day_Quinzena" = Table.SelectRows("#Expanded All Rows", each [Date] = [Max
Date by Day and Quinzena] ),

    #"Removed Other Columns1" = Table.SelectColumns("#Filtro Max
Day_Quinzena",{"Quinzena", "MesAno", "Date", "Version", "Hora", "Description"}),

    #"Changed Type1" = Table.TransformColumnTypes("#Removed Other
Columns1",{"Quinzena", type text}),

    #"Added Custom" = Table.AddColumn("#Changed Type1", "View_date", each [MesAno] &"-
"&[Quinzena])

in

    #"Added Custom"

```

```

// Etapas Preços CCEE #parte 1

let

    // Conexão e seleção da tabela

    Source = AmazonRedshift.Database(

        "your-cluster-endpoint.amazonaws.com",

        "your-database-name",

        [

            CredentialConnectionString = [

                UserName = "your-username",

```

```
// Etapas Preços CCEE #parte 2

    Password = "your-password",
        Encrypted = false,
        Host = "your-cluster-endpoint.amazonaws.com",
        Database = "your-database-name",
        Port = "5439"
    ]
]
),
Schema = Source([Name="your-schema-name",Kind="Schema"])[Data],
Table = Schema([Name="your-table-name",Kind="Table"])[Data],

//Fim Genérico Extração e navegação

#"Promoted Headers" = Table.PromoteHeaders(Table, [PromoteAllScalars=true]),
#"Filtered Rows" = Table.SelectRows("#Promoted Headers", each true),
#"Changed Type" = Table.TransformColumnTypes("#Filtered Rows",{{"Data", type datetime},
{"Price PLD", type number}, {"Submercado", type text}}),
#"Split Column by Delimiter" = Table.SplitColumn(Table.TransformColumnTypes("#Changed
Type", {"Data", type text}, "pt-BR"), "Data", Splitter.SplitTextByDelimiter(" ", QuoteStyle.Csv),
{"Date", "Hora"}),
#"Changed Type1" = Table.TransformColumnTypes("#Split Column by Delimiter",{{"Date",
type date}, {"Hora", type time}}),
#"Added Custom" = Table.AddColumn("#Changed Type1", "Month", each
Date.Month([Date])),
#"Grouped Rows" = Table.Group("#Added Custom", {"Date", "Month"}, {"Count", each
Table.RowCount(_, Int64.Type), {"All Rows", each _, type table [Date=nullable date,
Hora=nullable time, Price PLD=nullable number, Submercado=nullable text, Month=number]},
{"Avg Price PLD", each List.Average([Price PLD]), type nullable number}}),
#"Removed Other Columns" = Table.SelectColumns("#Grouped Rows",{"Date", "Avg Price
PLD"})
in
#"Removed Other Columns"
```

```
//Etapas Balanço Unificado #parte1
let
    // Conexão e seleção da tabela
    Source = AmazonRedshift.Database(
        "your-cluster-endpoint.amazonaws.com",
        "your-database-name",
        [
            CredentialConnectionString = [
                UserName = "your-username",
                Password = "your-password",
                Encrypted = false,
                Host = "your-cluster-endpoint.amazonaws.com",
                Database = "your-database-name",
                Port = "5439"
            ]
        ]
    ),
    Schema = Source([Name="your-schema-name",Kind="Schema"])[Data],
    Table = Schema([Name="your-table-name",Kind="Table"])[Data],
    //Fim Genérico Extração e navegação
    #"Promoted Headers" = Table.PromoteHeaders(Table, [PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"Entity", type text},
{"Activity", type text}, {"Type", type text}, {"Techno", type text}, {"Agrupamento", type text},
{"Date", type date}, {"Volume", type number}, {"Price", type number}, {"Version", Int64.Type}}),
    #"Merged Queries" = Table.NestedJoin(#"Changed Type", {"Version"},
balanco_Unificado_Versoes, {"Version"}, "Datamart EBE Version", JoinKind.LeftOuter),
    #"Expanded Datamart EBE Version" = Table.ExpandTableColumn(#"Merged Queries",
"Datamart EBE Version", {"Quinzena", "MesAno", "View_date"}, {"Datamart EBE
Version.Quinzena", "Datamart EBE Version.MesAno", "Datamart EBE Version.View_date"}),
    #"Changed Type1" = Table.TransformColumnTypes(#"Expanded Datamart EBE
Version",{{"Datamart EBE Version.Quinzena", type text}, {"Datamart EBE Version.MesAno",
type text}, {"Datamart EBE Version.View_date", type text}})
in
    #"Changed Type1"
```

```
//Etapas Preços Futuros #parte 1
let
    Source = SharePoint.Files("Pasta no Sharepoint", [ApiVersion = 15]),
    // Extração e navegação
    #"Filtered Hidden Files1" = Table.SelectRows(Source, each [Attributes]?[Hidden]? <> true),
    #"Invoke Custom Function1" = Table.AddColumn(#"Filtered Hidden Files1", "Transform File",
    each #"Transform File"([Content])),
    #"Renamed Columns1" = Table.RenameColumns(#"Invoke Custom Function1", {"Name",
    "Source.Name"}),
    #"Removed Other Columns1" = Table.SelectColumns(#"Renamed Columns1",
    {"Source.Name", "Transform File"}),
    #"Expanded Table Column1" = Table.ExpandTableColumn(#"Removed Other Columns1",
    "Transform File", Table.ColumnNames(#"Transform File"("#Sample File"))),
    #"Replaced Value2" = Table.ReplaceValue(#"Expanded Table
    Column1", ".", ",", Replacer.ReplaceText,{"Price"}),
    #"Changed Type" = Table.TransformColumnTypes(#"Replaced Value2",{{"Source.Name", type
    text}, {"View_date", type text}, {"Date", type date}, {"Price", type number}}),
    #"Replaced Value" = Table.ReplaceValue(#"Changed
    Type", ".csv", "", Replacer.ReplaceText,{"Source.Name"}),
    #"Replaced Value1" = Table.ReplaceValue(#"Replaced Value", "-
    ", "/", Replacer.ReplaceText,{"Source.Name"}),
    #"Renamed Columns" = Table.RenameColumns(#"Replaced Value1",{{"View_date",
    "MesAno"}})
in
    #"Renamed Columns"
```

```
// Etapas Entity Mapping
let
    // Referência à tabela original
    TabelaOriginal = balanço_Unificado,

    #"Grouped Rows" = Table.Group(TabelaOriginal, {"Entity"}, {"Count", each
Table.RowCount(_, Int64.Type)}),

    #"Removed Columns" = Table.RemoveColumns(#"Grouped Rows",{"Count"}),

    #"Added Conditional Column" = Table.AddColumn(#"Removed Columns", "Custom", each if
[Entity] = "EBE" then 1 else if [Entity] = "Geramamore" then 2 else if [Entity] = "Jirau" then 3
else 9)

in
#"Added Conditional Column"
```

```
// Etapas View_date Mapping #parte 1
let
    Source = SharePoint.Files("Pasta no Sharepoint", [ApiVersion = 15]),

    // Extração e navegação

    #"Filtered Hidden Files1" = Table.SelectRows(Source, each [Attributes]?[Hidden]? <> true),

    #"Invoke Custom Function1" = Table.AddColumn(#"Filtered Hidden Files1", "Transform File",
each #"Transform File"([Content])),

    #"Removed Other Columns" = Table.SelectColumns(#"Invoke Custom Function1",{"Date
created", "Name"}),

    #"Replaced Value3" = Table.ReplaceValue(#"Removed Other
Columns", ".csv", "", Replacer.ReplaceText, {"Name"}),

    #"Replaced Value4" = Table.ReplaceValue(#"Replaced Value3", "-
", "/" , Replacer.ReplaceText, {"Name"}),

    #"Added Suffix" = Table.TransformColumns(#"Replaced Value4", {"Name", each _ & "-1",
type text}),

    #"Added Index" = Table.AddIndexColumn(#"Added Suffix", "Index", 0, 10, Int64.Type),

    Custom1 = #"Added Suffix",

    #"Replaced Value" = Table.ReplaceValue(Custom1, "-1", "-2", Replacer.ReplaceText, {"Name"}),

    #"Added Index1" = Table.AddIndexColumn(#"Replaced Value", "Index", 1, 5, Int64.Type),

    #"Appended Query" = Table.Combine({"Added Index1", #"Added Index"}),

    #"Sorted Rows" = Table.Sort(#"Appended Query", {"Index", Order.Ascending}),
```

```
// Etapas View_date Mapping #parte 2

#"Renamed Columns" = Table.RenameColumns("#Sorted Rows",{"Name", "View_date"}),

#"Duplicated Column" = Table.DuplicateColumn("#Renamed Columns", "View_date",
"View_date - Copy"),

#"Split Column by Delimiter" = Table.SplitColumn("#Duplicated Column", "View_date -
Copy", Splitter.SplitTextByDelimiter("-", QuoteStyle.Csv), {"MesAno", "Quinzena"}),

#"Changed Type" = Table.TransformColumnTypes("#Split Column by Delimiter",{"MesAno",
type text}, {"Quinzena", Int64.Type})

in

#"Changed Type"
```

```
//Etapas dCalendar #parte 1

let

// Defina o início e o fim do intervalo de datas

DataInicio = #date(2018, 1, 1),

DataFim = #date(2024, 12, 31),

// Crie uma lista de datas entre DataInicio e DataFim

Dias = List.Dates(DataInicio, Duration.Days(Duration.From(DataFim - DataInicio)) + 1,
#duration(1, 0, 0, 0)),

// Converta a lista em uma tabela

TabelaDias = Table.FromList(Dias, Splitter.SplitByNothing(), {"Date"}),

// Adicione colunas adicionais para análise de tempo

#"Ano" = Table.AddColumn(TabelaDias, "Ano", each Date.Year([Date]), Int64.Type),

#"Mês" = Table.AddColumn("#Ano", "Mês", each Date.Month([Date]), Int64.Type),

#"Dia" = Table.AddColumn("#Mês", "Dia", each Date.Day([Date]), Int64.Type),

#"AnoMês" = Table.AddColumn("#Dia", "Ano-Mês", each Date.ToText([Date], "yyyy-MM")),

#"Nome do Mês" = Table.AddColumn("#AnoMês", "Nome do Mês", each Date.ToText([Date],
"MMMM"), type text),

#"Trimestre" = Table.AddColumn("#Nome do Mês", "Trimestre", each "Q" &
Number.ToText(Date.QuarterOfYear([Date]))),
```

```
//Etapas dCalendar #parte 2

#"Dia da Semana" = Table.AddColumn("#Trimestre", "Dia da Semana", each
Date.DayOfWeek([Date])),

#"Nome do Dia da Semana" = Table.AddColumn("#Dia da Semana", "Nome do Dia da
Semana", each Date.ToText([Date], "dddd"), type text),

#"Semana do Ano" = Table.AddColumn("#Nome do Dia da Semana", "Semana do Ano", each
Date.WeekOfYear([Date]), Int64.Type),

#"Changed Type" = Table.TransformColumnTypes("#Semana do Ano",{{"Date", type date}})

in

#"Changed Type"
```