

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
DE MINAS GERAIS (IFMG)
CAMPUS BAMBUÍ
BACHARELADO EM ENGENHARIA DE COMPUTAÇÃO

Vinícius Tadeu Andrade Costa

**ANÁLISE E INTEGRAÇÃO DE DADOS DO MERCADO FINANCEIRO BRASILEIRO
DISPONIBILIZADOS PELA B3 E CVM**

VINÍCIUS TADEU ANDRADE COSTA

**ANÁLISE E INTEGRAÇÃO DE DADOS DO MERCADO FINANCEIRO BRASILEIRO
DISPONIBILIZADOS PELA B3 E CVM**

Trabalho de conclusão de curso apresentado ao Curso de Bacharelado em Engenharia de Computação do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG) – *Campus* Bambuí para obtenção do grau de Bacharel em Engenharia de Computação.

Orientador: Marcos Roberto Ribeiro

Vinícius Tadeu Andrade Costa

**ANÁLISE E INTEGRAÇÃO DE DADOS DO MERCADO FINANCEIRO BRASILEIRO
DISPONIBILIZADOS PELA B3 E CVM**

Trabalho de conclusão de curso apresentado ao Curso de Bacharelado em Engenharia de Computação do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG) – *Campus* Bambuí para obtenção do grau de Bacharel em Engenharia de Computação.

Aprovado em 18 de Janeiro de 2025 pela banca examinadora:

Marcos Roberto Ribeiro – IFMG – *Campus* Bambuí – (Orientador)

Dedico este trabalho à minha esposa e filhos, incentivadores e fontes inesgotáveis de apoio, amor e compreensão.

AGRADECIMENTOS

Agradeço a toda à minha família, esposa, filhos, pais e minha irmã, por acreditarem em mim e pelo incentivo constante na realização deste trabalho.

Agradeço à minha orientadora, ao meu coorientador e a todos que contribuíram de alguma forma para a realização deste trabalho.

“A tarefa mais importante de uma pessoa que vem ao mundo é criar algo.”
Paulo Freire

RESUMO

Este trabalho é um modelo em \LaTeX utilizando a classe $\text{IFT}_{\text{E}}\text{X}2024$. Tal classe foi desenvolvida com base no manual de normalização de trabalhos acadêmicos do IFMG e nas normas relacionadas da Associação Brasileira de Normas Técnicas. Este modelo apresenta uma estrutura básica com exemplos de elementos pré e pós-textuais. Maiores informações sobre como utilizar a classe podem ser encontradas no manual da classe $\text{IFT}_{\text{E}}\text{X}2024$.

Palavras-chave: $\text{IFT}_{\text{E}}\text{X}2024$. Modelo. IFMG. \LaTeX .

ABSTRACT

This work is a template in \LaTeX using the `IFT \LaTeX 2024class`. This class was developed based on the academic work standardization manual of IFMG and the related norms of the Brazilian Association of Technical Standards. This template presents a basic structure with examples of pre and post-textual elements. Further information on how to use the class can be found in the `IFT \LaTeX 2024class` manual.

Keywords: `IFT \LaTeX 2024`. Template. IFMG. \LaTeX .

LISTA DE FIGURAS

| | |
|------------------------------------------------------------------------------|----|
| Figura 1 - Esquema lógico do banco de dados para integração de dados da CVM. | 24 |
| Figura 2 - Fluxograma do projeto | 32 |
| Figura 3 - Estrutura do FCA | 38 |

LISTA DE QUADROS

| | |
|--------------------------------------------------------------------|----|
| Quadro 1 - Especificações do computador utilizado no TCC | 33 |
| Quadro 2 - Comparativo entre conjuntos de dados da CVM | 45 |

LISTA DE TABELAS

| | |
|-------------------------------------------------------------------------------------------|----|
| Tabela 1 - Resumo dos campos da base Informação Cadastral | 53 |
| Tabela 2 - Arquivos Estruturados do Formulário Cadastral (FCA) | 56 |
| Tabela 3 - Documentos Estruturados das Informações Periódicas e Eventuais (IPE) | 58 |
| Tabela 4 - Arquivos Estruturados do FRE | 60 |

LISTA DE SIGLAS

IFMG – Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais

TCC – Trabalho de conclusão de curso

B3 – Brasil, Bolsa e Balcão

CVM – Comissão de Valores Mobiliários

DFP – Demonstrações Financeiras Padronizadas

ITR – Formulário de Informações Trimestrais

IPE – Periódicos e Eventuais

FRE – Formulário de Referência

LISTA DE SÍMBOLOS

\mathbb{X} – Variável X

\mathbb{R} – Conjunto dos números reais

SUMÁRIO

| | | |
|--------------|-------------------------------------------------------------------|------------------|
| 1 | INTRODUÇÃO | 15 |
| 1.1 | Justificativa | 16 |
| 1.2 | Objetivos | 16 |
| 1.3 | Resultados Esperados | 17 |
| 2 | FUNDAMENTOS TEÓRICOS | 18 |
| 2.1 | Mercado financeiro | 18 |
| 2.2 | Análises para investimento | 20 |
| 2.2.1 | <i>Indicadores de rentabilidade</i> | <i>21</i> |
| 2.2.2 | <i>Indicadores de Liquidez</i> | <i>22</i> |
| 2.2.3 | <i>Indicadores de Endividamento</i> | <i>22</i> |
| 2.2.4 | <i>Valor intrínseco</i> | <i>23</i> |
| 2.3 | Modelagem e integração de dados | 23 |
| 2.3.1 | <i>Modelagem dos dados e estruturação lógica</i> | <i>23</i> |
| 2.3.2 | <i>Processo de integração de dados</i> | <i>25</i> |
| 2.4 | Estado da arte | 26 |
| 2.4.1 | <i>Trabalhos acadêmicos</i> | <i>27</i> |
| 2.4.2 | <i>Projetos e plataformas similares</i> | <i>27</i> |
| 2.4.3 | <i>Diferencial do trabalho proposto</i> | <i>28</i> |
| 3 | METODOLOGIA | 30 |
| 3.1 | Classificação da Pesquisa | 30 |
| 3.2 | Gerenciamento do Projeto | 31 |
| 3.3 | Solução proposta | 31 |

| | | |
|--------------|-------------------------------------------------------------------------------|-----------|
| 3.4 | Materiais e Tecnologias | 32 |
| 4 | DESENVOLVIMENTO | 35 |
| 4.1 | Análise inicial dos dados da CVM | 35 |
| 4.1.1 | <i>Informação cadastral</i> | 37 |
| 4.1.2 | <i>Formulário cadastral</i> | 37 |
| 4.1.3 | <i>Informações periódicas e eventuais</i> | 38 |
| 4.1.4 | <i>Formulário de Referência</i> | 40 |
| 4.1.5 | <i>Valores mobiliários negociados e detidos</i> | 41 |
| 4.1.6 | <i>Demonstrativos financeiros padronizados</i> | 42 |
| 4.1.6.1 | Informações trimestrais | 43 |
| 4.1.6.2 | Formulário de demonstrações financeiras padronizadas | 43 |
| 4.1.7 | <i>Informe do Código de Governança</i> | 43 |
| 4.1.8 | <i>Resumo dos dados iniciais da CVM</i> | 45 |
| 4.2 | Modelagem e Estruturação dos Dados | 46 |
| 4.3 | Características e Benefícios Analíticos da Estrutura Final | 47 |
| 4.4 | Software | 48 |
| 5 | CONCLUSÃO | 49 |
| | REFERÊNCIAS | 50 |
| | GLOSSÁRIO | 51 |
| | APÊNDICES | 52 |
| | APÊNDICE A - Dicionário Resumido — Informação Cadastral | 53 |
| | APÊNDICE B - Resumo do Formulário Cadastral | 56 |
| | APÊNDICE C - Lista de documentos disponíveis no conjunto IPE | 58 |
| | APÊNDICE D - Arquivos estruturados do FRE | 60 |

| | |
|--------------------------------------------------------------------|-----------|
| APÊNDICE E - Código para baixar os dados da CVM | 63 |
| APÊNDICE F - Código para extrair os dados da CVM | 66 |
| APÊNDICE G - Mapeamento Completo dos Dados da CVM | 68 |
| APÊNDICE H - Primeira modelagem | 75 |
| APÊNDICE I - Modelagem apos ajustes iniciais | 77 |
| APÊNDICE J - Penúltima modelagem | 79 |
| APÊNDICE K - Modelagem presente no código | 81 |
| ÍNDICE | 83 |

1 INTRODUÇÃO

O mercado financeiro brasileiro tem passado por transformações significativas impulsionadas pelo avanço tecnológico e pela crescente demanda por soluções que automatizem a integração e a análise de dados (**dantas:2020:comportamento**). Esse movimento reflete a busca constante por maior eficiência e acessibilidade nas informações financeiras, fundamentais para investidores e analistas que utilizam a análise fundamentalista como base para a tomada de decisões.

O crescimento desse interesse pode ser observado no relatório anual de 2023 da Brasil, Bolsa e Balcão (B3), que apontou um aumento de aproximadamente 15% no número de investidores em comparação com 2022 (**b3:2023:relatorio**). Esse cenário reforça a necessidade de garantir o acesso simplificado e estruturado às informações financeiras das empresas, promovendo maior eficiência nas análises e decisões de mercado.

No Brasil, dois órgãos desempenham papéis centrais na disponibilização de dados do mercado de capitais: a B3 e a Comissão de Valores Mobiliários (CVM). A B3, fundada em 1890 e sediada em São Paulo, é a única bolsa de valores do país e fornece dados sobre o histórico de negociações de ativos (**b3:2023:investidores**).

Já a CVM, criada em 1976, é a entidade responsável por regulamentar e supervisionar o mercado de capitais brasileiro, disponibilizando publicamente uma ampla gama de informações contábeis e financeiras das empresas listadas na bolsa (**cvm:2009:informacao**). Essas informações são essenciais para investidores que utilizam a análise fundamentalista, metodologia que avalia a saúde financeira e o desempenho das empresas com base em seus demonstrativos financeiros e outros indicadores econômicos.

Embora a CVM disponibilize esses dados publicamente, seu formato dificulta a manipulação por usuários sem conhecimento técnico avançado. Muitos estão em arquivos complexos que exigem processamento adicional, criando um obstáculo para investidores e analistas que necessitam de informações ágeis e acessíveis. Essa barreira técnica restringe o acesso à informação e limita a capacidade analítica de boa parte do mercado.

A necessidade de soluções mais acessíveis já foi abordada por estudos acadêmicos. (**deAraujo:2021:modeloDados**) propõem um modelo de dados flexível para análise fundamentalista moderna, estruturando balanço patrimonial, demonstração de resultados e fluxo de caixa em um banco de dados não relacional (MongoDB). Esse tipo de abordagem reforça a importância de alternativas que simplifiquem o acesso e a análise dos dados financeiros.

Diante desse contexto, este trabalho propõe o desenvolvimento de um sistema automatizado para a integração e análise dos dados fornecidos pela CVM, facili-

tando seu acesso e processamento. A solução busca oferecer uma ferramenta capaz de extrair e estruturar essas informações de forma eficiente, contribuindo para a análise fundamentalista e auxiliando na tomada de decisões. Além disso, pretende otimizar o processo de análise, tornando-o mais acessível e organizado para investidores, analistas e pesquisadores. Com essa abordagem, espera-se reduzir a complexidade no manuseio dos dados e aprimorar a compreensão do mercado.

1.1 Justificativa

A B3 apresentou um crescimento significativo no volume de negócios e de capital movimentado ao longo dos anos. Em 2023, o volume total negociado na B3 alcançou R\$ 7,2 trilhões, representando um aumento de 12,6% em relação a 2022 (**b3:2023:investidores**). Além disso, o montante total de dinheiro movimentado na B3 em 2023 foi de R\$ 2,4 trilhões, um aumento de 15,5% em relação ao ano anterior (**b3:2023:investidores**).

Embora os dados da B3 não estejam diretamente presentes neste trabalho, seu contexto é fundamental, uma vez que ela representa o principal ambiente de negociação de ativos no Brasil. As informações que embasam os negócios realizados na B3, especialmente aquelas relacionadas às demonstrações financeiras e dados cadastrais das empresas listadas, estão disponíveis na CVM. Dessa forma, a integração dos dados da CVM serve como uma base fundamental para análises que indiretamente impactam o entendimento do mercado da B3.

Atualmente, as informações da CVM estão disponíveis em formatos específicos, fragmentados e sem integração adequada, o que dificulta o acesso e a análise automatizada dos dados. Até o momento, não foram encontrados trabalhos que promovam essa integração de forma estruturada e com disponibilização pública.

A proposta de integrar os dados da CVM visa criar uma base de dados unificada e acessível, que sirva como suporte para futuros trabalhos acadêmicos e aplicações práticas em algoritmos de análise de mercado financeiro. Com isso, espera-se facilitar o desenvolvimento de estudos e ferramentas que contribuam para uma melhor compreensão do mercado de capitais brasileiro (**lindman:2020:integration**)

1.2 Objetivos

Este estudo tem como objetivo principal o desenvolvimento de um sistema automatizado capaz de integrar e facilitar o acesso aos dados financeiros disponibilizados publicamente pela CVM. A proposta busca contribuir com a análise fundamentalista e apoiar o processo de tomada de decisão por parte de investidores, pesquisadores e demais interessados no mercado financeiro.

Para atingir esse objetivo, são estabelecidas as seguintes metas específicas:

- analisar a estrutura dos dados públicos fornecidos pela CVM, identificando suas fontes, formatos e padrões;
- propor uma modelagem lógica que permita a integração eficiente e automatizada desses dados;
- projetar e implementar um sistema que realize a extração, o processamento e a disponibilização dos dados de forma acessível e estruturada.

1.3 Resultados Esperados

Espera-se que a ferramenta desenvolvida neste trabalho integre de forma eficiente os dados financeiros públicos disponibilizados pela CVM, permitindo a criação de uma base de dados unificada, estruturada e acessível.

Essa base será construída com foco em um recorte histórico que contemple, no mínimo, o período a partir de 2019, o que proporcionará maior profundidade nas análises e permitirá identificar padrões e variações ao longo do tempo. Tal histórico é essencial para estudos mais robustos e comparativos dentro do contexto da análise fundamentalista.

Embora os dados da B3 não sejam utilizados diretamente neste trabalho, o sistema proposto pode contribuir indiretamente para análises relacionadas à bolsa de valores, uma vez que as informações integradas da CVM servem como base para decisões tomadas por investidores e instituições que atuam no ambiente da B3.

A disponibilização pública dessa base de dados tem o potencial de promover maior transparência, acessibilidade e democratização das informações financeiras no Brasil, beneficiando pesquisadores, investidores e demais interessados no mercado de capitais.

2 FUNDAMENTOS TEÓRICOS

Este capítulo apresenta os fundamentos teóricos que sustentam o desenvolvimento deste trabalho, bem como os principais estudos e iniciativas relacionadas ao tema. A Seção 2.1 aborda o mercado financeiro brasileiro, com ênfase na atuação da CVM e sua função reguladora. A Seção 2.2 discute as diferentes abordagens de análise para investimentos, com foco na análise fundamentalista. A Seção 2.3 descreve a modelagem dos dados e os processos de integração utilizados no projeto. Por fim, a Seção 2.4 apresenta trabalhos acadêmicos e iniciativas similares à proposta deste estudo, destacando seus principais diferenciais.

2.1 Mercado financeiro

O mercado financeiro é um sistema que engloba uma variedade de instituições, instrumentos e atividades relacionadas à gestão de recursos financeiros. Ele abrange diferentes segmentos, incluindo o mercado de capitais, o mercado monetário, o mercado cambial e o mercado de derivativos. O principal propósito do mercado financeiro é facilitar a alocação eficiente de recursos entre poupadores e investidores, fornecendo oportunidades de investimento e financiamento para empresas, governos e indivíduos (**teixeira:2019:mercado; damodaran:2012:investimentos**).

No Brasil, o mercado financeiro é composto pelos segmentos de mercado de capitais, mercado monetário, mercado de crédito, mercado de câmbio e mercado de derivativos. O mercado de capitais envolve a negociação de ações, debêntures e outros ativos mobiliários. Esse segmento é fundamental para o financiamento das empresas, permitindo que elas captem recursos por meio da emissão de títulos negociáveis no mercado financeiro (**reis:2021:negociacao**).

O mercado monetário é o espaço onde ocorrem operações de curto prazo entre instituições financeiras e o Banco Central. Essas operações têm como objetivo controlar a liquidez da economia e a taxa de juros, garantindo o equilíbrio do sistema financeiro (**deSouzaFigueiredo:2021:mercado**).

O mercado de crédito destina-se ao financiamento de empresas e pessoas físicas por meio de empréstimos e financiamentos bancários. Esse segmento é essencial para a dinamização da economia, permitindo investimentos e consumo a partir da concessão de crédito.

O mercado de câmbio é responsável pelas transações de compra e venda de moedas estrangeiras. Sua função principal é viabilizar o comércio internacional, investimentos estrangeiros e a gestão de reservas cambiais (**gois:2019:efeito**).

O mercado de derivativos compreende a negociação de contratos futuros, opções e *swaps*, instrumentos utilizados para a gestão de riscos financeiros. Esse

segmento permite que investidores e empresas se protejam contra variações adversas nos preços de ativos e taxas de juros (**figueiredo:2023:capacidade**).

Dentro do mercado financeiro, as ações representam uma das principais modalidades de investimento. Uma ação é uma fração do capital social de uma empresa, conferindo ao seu detentor a propriedade de uma parcela da empresa emissora. Ao adquirir ações, o investidor se torna um sócio da empresa e pode ter direito a participação nos lucros distribuídos e, em alguns casos, a voto nas assembleias de acionistas (**reis:2021:negociacao; gomes:2007:bolsa**).

A negociação de ações ocorre principalmente em bolsas de valores, como a B3 no Brasil, onde investidores podem comprar e vender ativos. O preço das ações é determinado pela oferta e demanda, ou seja, pela interação entre compradores e vendedores no mercado (**santander:2024:pregao**). Além disso, diversos fatores influenciam a valorização ou desvalorização dos papéis, como os resultados financeiros das empresas, a conjuntura econômica e eventos geopolíticos (**damodaran:2012:investimentos; attie:2013:bolsa**).

Duas instituições fundamentais nesse ecossistema são a B3 e a CVM. A B3 atua como principal plataforma de negociação de ativos financeiros, enquanto a CVM desempenha um papel regulador crucial, garantindo transparência e equidade no mercado de capitais (**cvm:2023:funcoes**).

Criada pela Lei nº 6.385/1976 (**brasil:1976:lei6385**), a CVM tem como principal missão proteger os investidores e garantir o funcionamento eficiente do mercado de capitais (**cvm:2009:informacao**). Entre suas principais atribuições estão (**cvm:2009:informacao; roberto:2023:papel**):

- regulamentação da abertura de capital de empresas e suas obrigações com o mercado;
- supervisão de corretoras, bancos de investimento e outras instituições financeiras;
- fiscalização da negociação de ativos financeiros e combate a fraudes e manipulações de mercado;
- garantia da transparência e divulgação de informações pelas companhias de capital aberto.

A CVM exige que todas as empresas listadas na bolsa de valores publiquem periodicamente suas demonstrações financeiras e outros documentos essenciais, como balanços patrimoniais e informações sobre acionistas relevantes. Essa transparência fortalece a confiança dos investidores e contribui para a estabilidade e o crescimento do mercado financeiro brasileiro (**fgv:2024:transformacao**).

Dessa forma, a regulamentação e a supervisão do mercado financeiro brasileiro desempenham um papel essencial na proteção dos investidores e no desenvolvimento sustentável do setor (**figueiredo:2023:capacidade**).

Para garantir a ampla disponibilidade dessas informações, a CVM mantém uma base de dados acessível ao público por meio de seu portal eletrônico¹ e sistemas especializados, como o Sistema Empresas.NET e o CVMWeb. As empresas registradas são obrigadas a enviar seus documentos periodicamente, os quais são disponibilizados em diversos formatos, como PDF, XML e XBRL, permitindo tanto a consulta manual quanto a análise automatizada (**cvm:2009:informacao**). Essa estrutura facilita o acompanhamento do mercado por investidores, analistas e demais interessados, promovendo maior transparência e acesso à informação.

A base de dados da CVM inclui:

- demonstrações financeiras padronizadas (DFPs);
- informes trimestrais (ITRs);
- relatórios de governança corporativa;
- documentos de ofertas públicas de ações (IPO);
- dados sobre fundos de investimento e gestores de ativos.

2.2 Análises para investimento

A análise de investimentos busca avaliar oportunidades de alocação de capital, utilizando diferentes abordagens para identificar ativos com potencial de retorno (**liaw:2011:business**). Entre essas abordagens, destacam-se a análise técnica, a análise de sentimento, a análise quantitativa, a análise macroeconômica e a análise fundamentalista.

A análise técnica baseia-se no estudo de padrões gráficos e estatísticos do histórico de preços dos ativos. Os analistas técnicos utilizam indicadores como médias móveis, bandas de Bollinger e índices de força relativa para identificar tendências e pontos de compra ou venda no mercado (**omane:2019:time**).

A análise de sentimento considera o impacto de notícias, redes sociais e opiniões de investidores sobre os ativos financeiros. Essa abordagem busca medir o comportamento do mercado a partir da análise de textos, discursos e volumes de menções relacionadas a determinados ativos, utilizando técnicas de processamento de linguagem natural e aprendizado de máquina (**kearney:2014:textual**).

A análise quantitativa utiliza modelos matemáticos e estatísticos para prever movimentos de mercado e identificar oportunidades de investimento. Esse método frequentemente emprega algoritmos computacionais, regressões estatísticas e aprendizado de máquina para encontrar correlações e padrões em grandes volumes de dados financeiros (**sahu:2023:overview**).

A análise macroeconômica avalia fatores econômicos amplos, como inflação, Produto Interno Bruto (PIB), taxas de juros e política monetária. Essa aborda-

¹ <https://dados.cvm.gov.br>.

gem busca compreender como variáveis macroeconômicas influenciam os mercados financeiros e os preços dos ativos, sendo fundamental para investidores institucionais e gestores de fundos (**claessens:2017:macroeconomic**).

A análise fundamentalista foca nos fundamentos financeiros das empresas, examinando indicadores como receita, lucro, endividamento e governança corporativa. Esse método busca determinar o valor intrínseco de um ativo, comparando-o ao seu preço de mercado para identificar se está subvalorizado ou supervalorizado. No contexto deste estudo, a análise fundamentalista será o enfoque principal, dada sua relevância na avaliação detalhada do desempenho financeiro e estratégico das companhias.

A análise fundamentalista investiga os dados financeiros e operacionais das empresas para determinar seu valor intrínseco (**mathew:2024:overview**). Para isso, utiliza diversos indicadores que auxiliam na avaliação do desempenho financeiro, da estrutura de capital e da liquidez da empresa. Esses indicadores podem ser agrupados em quatro categorias principais: indicadores de rentabilidade, indicadores de liquidez, indicadores de endividamento e valor intrínseco.

2.2.1 Indicadores de rentabilidade

Os indicadores de rentabilidade medem a eficiência da empresa em gerar lucros em relação a diferentes bases financeiras. Esses cálculos são amplamente utilizados em estudos financeiros e derivam da contabilidade gerencial e da análise de demonstrações financeiras (**kothari:2001:capital**).

O lucro por ação (LPA) indica o montante de lucro líquido atribuído a cada ação ordinária em circulação da empresa. É um indicador essencial para avaliar o desempenho das ações no mercado. A fórmula é dada pela Equação (2.1a). Considerando um lucro líquido de R\$ 10 milhões e 2 milhões de ações em circulação, tem-se o LPA de R\$ 5,00, conforme a Equação (2.1b) (**omane:2019:time**).

$$\text{LPA} = \frac{\text{Lucro líquido}}{\text{Número de ações}} \quad (2.1a) \quad \text{LPA} = \frac{10.000.000,00}{2.000.000,00} = 5,00 \quad (2.1b)$$

O índice preço por lucro (P/L) mede quanto os investidores estão dispostos a pagar pelo lucro da empresa. A fórmula está na Equação (2.2a). Se uma ação custa R\$ 50,00 e o LPA é R\$ 5,00, o P/L será 10,00, indicando que os investidores estão dispostos a pagar 10 vezes o lucro anual por ação (**sahu:2023:overview**).

$$\text{P/L} = \frac{\text{Preço da ação}}{\text{LPA}} \quad (2.2a) \quad \text{P/L} = \frac{50,00}{5,00} = 10,00 \quad (2.2b)$$

O *dividend yield* (DY) mede a rentabilidade dos dividendos pagos em relação ao preço da ação. Se a empresa paga R\$ 2,00 de dividendos por ação e o

preço da ação é R\$ 40,00, o cálculo da Equação (2.3b) resulta em um DY de 5,00% (**kearney:2014:textual**).

$$DY = \frac{\text{Dividendo por ação}}{\text{Preço da ação}} \times 100 \quad (2.3a) \qquad DY = \frac{2,00}{40,00} \times 100 = 5,00\% \quad (2.3b)$$

2.2.2 Indicadores de Liquidez

Os indicadores de liquidez avaliam a capacidade da empresa de honrar suas obrigações de curto prazo (**claessens:2017:macroeconomic**). A Liquidez Corrente (LC) é expressa na Equação (2.4a). Se o ativo circulante é de R\$ 500.000 e o passivo circulante de R\$ 250.000, a liquidez corrente é igual a 2,00, conforme a Equação (2.4b).

$$LC = \frac{\text{Ativo circulante}}{\text{Passivo circulante}} \quad (2.4a) \qquad LC = \frac{500.000}{250.000} = 2,00 \quad (2.4b)$$

2.2.3 Indicadores de Endividamento

Os indicadores de endividamento mostram a dependência da empresa em relação ao capital de terceiros (**kothari:2001:capital**).

Um dos principais indicadores é a dívida bruta, que representa a soma do passivo circulante e do passivo não circulante da empresa, conforme definido na Equação (2.5a). A Equação (2.5a) expressa a fórmula geral da dívida bruta, que é composta pelos valores de obrigações de curto prazo (passivo circulante) e de longo prazo (passivo não circulante). Já a Equação (2.5b) apresenta um exemplo numérico, onde a dívida bruta totaliza R\$ 3.000.000, a partir da soma de R\$ 1.000.000 de passivo circulante e R\$ 2.000.000 de passivo não circulante.

$$\text{Dívida bruta} = \text{Passivo circulante} + \text{Passivo não circulante} \quad (2.5a)$$

$$\text{Dívida bruta} = 1.000.000 + 2.000.000 = 3.000.000 \quad (2.5b)$$

Por exemplo, se a dívida bruta é de R\$ 3.000.000 e o caixa e equivalentes de caixa totalizam R\$ 500.000, a dívida líquida será de R\$ 2.500.000, conforme indicado na Equação (2.6b). A Equação (2.6a) define a dívida líquida, que representa o valor da dívida efetiva da empresa, descontado o montante disponível em caixa ou em aplicações de liquidez imediata. Já a Equação (2.6b) ilustra esse conceito com

valores hipotéticos, demonstrando que, ao subtrair R\$ 500.000 disponíveis em caixa de uma dívida bruta de R\$ 3.000.000, obtém-se uma dívida líquida de R\$ 2.500.000.

$$\text{Dívida líquida} = \text{Dívida bruta} - \text{Caixa e equivalentes de caixa} \quad (2.6a)$$

$$\text{Dívida líquida} = 3.000.000 - 500.000 = 2.500.000 \quad (2.6b)$$

2.2.4 Valor intrínseco

O valor intrínseco é estimado pelo método de Desconto de Fluxo de Caixa (DCF), conforme a Equação (2.7a). Nessa equação, r representa a taxa de desconto e t o período de tempo. Com fluxo de caixa de R\$ 1.000.000 e taxa de desconto de 10%, o valor intrínseco resulta em R\$ 909.090,91, conforme demonstrado na Equação (2.7b). Esse cálculo reflete o valor presente do fluxo de caixa futuro (**figueiredo:2023:capacidade**).

$$\text{Valor intrínseco} = \sum \frac{\text{Fluxo de caixa esperado}}{(1 + r)^t} \quad (2.7a)$$

$$\text{Valor intrínseco} = \frac{1.000.000}{(1 + 0,10)^1} = \frac{1.000.000}{1,10} = 909.090,91 \quad (2.7b)$$

2.3 Modelagem e integração de dados

A modelagem e integração de dados são etapas fundamentais para a construção de um sistema automatizado de análise fundamentalista. Neste trabalho, essas etapas têm o propósito de estruturar, consolidar e organizar os dados disponibilizados pela CVM, garantindo sua utilização eficiente e integrada. A seguir, são apresentadas as principais abordagens empregadas na modelagem dos dados, bem como os métodos utilizados para sua integração, assegurando a consistência, acessibilidade e interoperabilidade das informações no sistema desenvolvido.

2.3.1 Modelagem dos dados e estruturação lógica

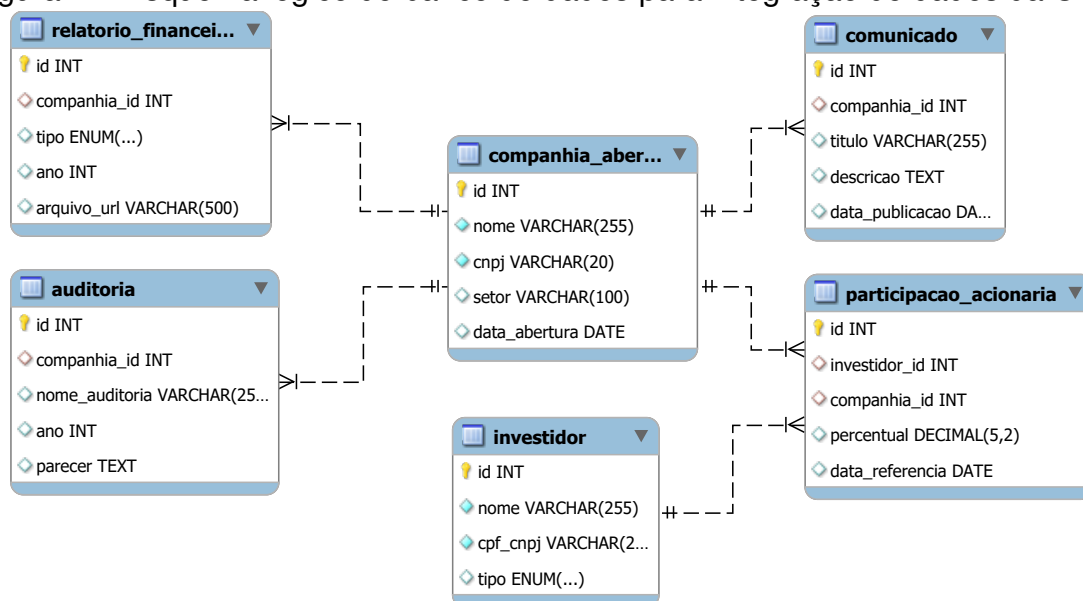
A modelagem de dados é uma etapa fundamental no desenvolvimento de sistemas que manipulam grandes volumes de informações, especialmente em contextos financeiros, onde a integridade, consistência e escalabilidade são requisitos indispensáveis (**elmasri:2005:sistemas**). Neste trabalho, adotou-se uma abordagem centrada na modelagem lógica, com o objetivo de estruturar uma base de dados rela-

cional robusta, capaz de integrar e armazenar dados públicos oriundos da CVM.

A modelagem lógica traduz os requisitos identificados no domínio do problema em uma estrutura formal que define, de maneira precisa, as tabelas, os atributos, os tipos de dados e as restrições de integridade necessárias para a representação das entidades envolvidas (**connolly:2015:database**). Diferente da modelagem conceitual, que geralmente se utiliza de representações como o Diagrama Entidade-Relacionamento (DER), a modelagem lógica já incorpora aspectos técnicos relacionados à implementação em um Sistema de Gerenciamento de Banco de Dados (SGBD), neste caso, o MySQL.

A Figura 1 apresenta o esquema lógico elaborado com base em práticas de engenharia reversa e análise das documentações e estruturas disponibilizadas pela CVM. Essa técnica permite extrair o modelo de dados a partir de fontes existentes, facilitando o redesenho e a normalização da base de dados (**scannapieco:2006:data**).

Figura 1 – Esquema lógico do banco de dados para integração de dados da CVM.



Fonte: Elaborado pelo autor, 2024.

O modelo proposto é composto pelas seguintes tabelas:

- *companhia_aberta*
- *comunicado*
- *investidor*
- *participacao_acionaria*
- *auditoria*
- *relatorio_financeiro*

A tabela *companhia_aberta* centraliza os dados cadastrais das empresas registradas na CVM, contendo atributos como *nome*, *cnpj*, *setor* e *data_abertura*. O campo *cnpj* possui uma restrição de unicidade (*UNIQUE*) para garantir a integridade

dos dados identificadores.

A tabela *comunicado* armazena registros de publicações oficiais vinculadas às companhias, sendo referenciada pela chave estrangeira *companhia_id*, que estabelece uma relação com a tabela *companhia_aberta*. Esse vínculo viabiliza a rastreabilidade entre os comunicados e as empresas emissores.

A tabela *investidor* registra os agentes que detêm participações acionárias, identificados pelo atributo *cpf_cnpj*, e categorizados segundo o tipo de investidor (*Pessoa Física* ou *Pessoa Jurídica*), utilizando a estrutura *ENUM* para controle de domínio.

A tabela *participacao_acionaria* representa a associação entre investidores e companhias abertas, permitindo o registro da proporção de ações detidas por data de referência. Essa tabela possui chaves estrangeiras para as tabelas *investidor* e *companhia_aberta*, o que possibilita a análise de composição societária ao longo do tempo.

A tabela *auditoria* documenta os pareceres emitidos por empresas auditoras sobre as demonstrações financeiras das companhias, detalhando o ano do parecer, o nome da auditoria e a descrição do conteúdo emitido.

Por fim, a tabela *relatorio_financeiro* concentra os metadados dos arquivos financeiros disponibilizados pela CVM, como os Demonstrativos Financeiros Padronizados (DFP), Informações Trimestrais (ITR), o Formulário de Referência e o IAN. Cada registro inclui o tipo do documento, o ano de referência e o link para o arquivo original.

Essa estrutura relacional foi projetada visando à normalização dos dados, conforme as boas práticas de projeto de bancos de dados relacionais (**elmasri:2016:fundamentals**). A modelagem lógica aqui apresentada oferece a base necessária para uma integração eficiente e escalável, compatível com análises automatizadas de caráter fundamentalista. Sua implementação favorece a consistência da base de dados, o desempenho nas consultas SQL e a reutilização de componentes em diferentes módulos do sistema.

2.3.2 Processo de integração de dados

A integração de dados é um procedimento essencial que consolida informações oriundas de diversas fontes, garantindo consistência, qualidade e acessibilidade para análises estratégicas e operacionais. Esse processo transforma dados brutos em informações valiosas para a tomada de decisão e é composto por etapas interdependentes: extração, transformação e carga.

Na etapa de extração, os dados são coletados diretamente de múltiplas fontes, como os repositórios da CVM. Essa fase utiliza técnicas que asseguram a captura completa e precisa das informações, minimizando riscos de perda ou corrupção. No presente trabalho, essa etapa foi implementada por meio de *scripts* automatiza-

dos em Python, que realizam o download de arquivos compactados disponibilizados pela CVM, como o `dfp_cia_aberta_2022.zip`. O sistema verifica a existência prévia dos arquivos e sua integridade antes de prosseguir com o processamento, evitando redundâncias e falhas.

Após a extração, os dados brutos passam por um processo de transformação que envolve padronização, limpeza, normalização e aplicação de regras de negócio. Essa etapa é responsável por eliminar inconsistências, corrigir valores ausentes, ajustar formatos e estruturar os dados para análises mais sofisticadas. No contexto deste projeto, foram realizadas conversões de tipos, unificação de formatos de datas e remoção de duplicidades. Informações como o código da conta contábil (`CD_CONTA`) foram associadas às suas descrições completas com base nos planos de contas divulgados pela CVM, garantindo maior legibilidade e utilidade analítica.

Na etapa final, os dados transformados são carregados em um banco de dados relacional estruturado, visando facilitar a manipulação, a consulta e a análise posterior. Para este trabalho, optou-se pelo uso do SQLite como solução leve e eficiente, adequada ao escopo local da ferramenta desenvolvida. As tabelas foram criadas respeitando normas de integridade referencial, com chaves primárias e estrangeiras bem definidas. Dados financeiros, cadastrais e operacionais foram organizados em tabelas como `demonstrativo_financeiro`, `informacao_trimestral` e `empresas`, possibilitando análises temporais, cálculo de indicadores como LPA e P/L, entre outros.

A integração harmônica dessas etapas resultou em uma base de dados confiável e atualizada, que sustenta a automatização dos processos analíticos e permite a construção de modelos quantitativos e preditivos. Essa abordagem se mostra fundamental para a aplicação da análise fundamentalista no contexto do mercado financeiro, proporcionando uma visão estruturada e estratégica para a tomada de decisão (**costa:2024:integracao; halevy:2006:data**).

2.4 Estado da arte

Nesta seção, são analisados os principais estudos acadêmicos e projetos práticos que fundamentam a abordagem deste trabalho. Inicialmente, apresenta-se uma síntese dos trabalhos existentes, destacando seus objetivos, metodologias empregadas e limitações encontradas. Em seguida, evidencia-se como o presente estudo propõe uma integração inédita de dados e técnicas, superando as deficiências das abordagens anteriores.

2.4.1 Trabalhos acadêmicos

Os estudos acadêmicos voltados à modelagem e à análise de dados financeiros têm se concentrado na automatização da análise fundamentalista e na adaptação de modelos tradicionais ao contexto do mercado brasileiro. Por exemplo, **montoia:2021:automatizacao** desenvolvem um sistema automatizado para a análise de ações, enfatizando a melhoria na rapidez e na eficiência da tomada de decisões. Esse trabalho detalha a aplicação de algoritmos de processamento de dados que reduzem a intervenção manual, permitindo uma análise em tempo real dos indicadores financeiros.

De forma complementar, **deAraujo:2021:modelo** propôs um modelo flexível que adapta os conceitos da análise fundamentalista para refletir as particularidades dos dados nacionais, considerando variáveis específicas do mercado brasileiro. Esse estudo destaca a importância da customização dos parâmetros e a integração de variáveis macroeconômicas na modelagem financeira.

Além disso, **delalibera:2023:automatizacao** aprimorou metodologias tradicionais, como o Modelo Rojo, para avaliação de ativos, enfatizando a precisão na quantificação de riscos e oportunidades de investimento. Este trabalho apresenta uma abordagem robusta que combina técnicas estatísticas com algoritmos de *machine learning* para melhorar a acurácia das previsões. Por sua vez, **reis:2020:analise** investigou a aplicação prática desses métodos na análise das ações negociadas, proporcionando *insights* sobre a volatilidade do mercado e a relevância de diferentes indicadores financeiros.

Outros estudos, como o de **vieira:2019:montagem**, analisam estratégias de montagem de carteiras de ações, comparando a performance das carteiras com índices de mercado e ressaltando as dificuldades na diversificação e no balanceamento dos ativos. Da mesma forma, **freitas:2020:analise** explorou a aplicação criteriosa de indicadores no setor bancário, evidenciando os desafios em medir a solidez financeira e a eficiência operacional de instituições financeiras.

2.4.2 Projetos e plataformas similares

No campo dos projetos práticos, diversas iniciativas presentes em repositórios, especialmente no GitHub, vêm se destacando por oferecer ferramentas e bases de dados voltadas para a extração e análise de informações financeiras. Por exemplo, o projeto de **paiva:2025:projetoGitHub** apresenta um sistema integrado que coleta, processa e organiza dados contábeis de empresas. Essa ferramenta tem como objetivo facilitar a aplicação da análise fundamentalista ao disponibilizar dados de forma estruturada e acessível, promovendo maior transparência e facilidade na análise dos

balanços.

De maneira similar, **minas:2025:projetoGitHub** propõe uma abordagem inovadora para a análise de balanços empresariais, enfatizando a importância da normalização dos dados e da criação de indicadores personalizados que se ajustem às especificidades de cada setor. Repositórios como os de **fontinele:2025:projetoGitHub** e **louredo:2025:projetoGitHub** se destacam ao desenvolver ferramentas específicas para interpretar demonstrações financeiras a partir de dados públicos, permitindo que o usuário identifique rapidamente pontos críticos e oportunidades de investimento. Contudo, grande parte dessas iniciativas se restringe à coleta e ao processamento dos dados, sem oferecer uma análise histórica contínua que possibilite a compreensão das tendências de longo prazo.

2.4.3 Diferencial do trabalho proposto

O diferencial do presente trabalho reside na integração de um histórico abrangente de dados financeiros com um sistema automatizado de acesso e análise das informações públicas disponibilizadas pela CVM. Embora os experimentos realizados tenham utilizado um recorte temporal de cinco anos, a ferramenta não se limita a esse intervalo. Enquanto a CVM mantiver a estrutura dos dados em seus repositórios, o sistema permanecerá funcional e poderá ser continuamente alimentado com novos dados, garantindo sua atualização e longevidade.

A proposta se distingue, ainda, por tratar-se de uma solução de código aberto, o que permite sua auditoria, personalização e extensão por outros pesquisadores e profissionais interessados. Essa abertura favorece a colaboração contínua, possibilitando o aprimoramento da ferramenta e sua adaptação a novos contextos analíticos ou setores econômicos específicos.

A abordagem adotada neste trabalho viabiliza tanto a avaliação contínua e comparativa dos dados quanto a análise contextualizada das informações. A avaliação contínua é sustentada pelo acúmulo de dados históricos, o que permite identificar tendências e padrões relevantes ao longo do tempo. Já a análise contextualizada emerge da capacidade de combinar dados históricos e atuais, ampliando a compreensão sobre ciclos econômicos, sazonalidades e eventos que impactam o mercado.

As principais inovações e contribuições do sistema desenvolvido podem ser resumidas nos seguintes pontos:

- integração de dados históricos financeiros em larga escala;
- processamento automatizado, com foco em escalabilidade e reusabilidade;
- contextualização dos indicadores contábeis e financeiros;
- manutenção da funcionalidade do sistema diante de novas atualizações

- da CVM, desde que preservada a estrutura de dados;
- disponibilização em formato de código aberto, fomentando a transparência e a colaboração.

Ao superar as limitações de iniciativas anteriores, que se restringem majoritariamente à coleta e estruturação pontual dos dados, a solução apresentada promove uma abordagem sistemática e extensível, capaz de subsidiar análises fundamentalistas robustas e atualizadas. Essa característica torna o sistema especialmente útil para aplicações acadêmicas, institucionais e profissionais voltadas à tomada de decisão baseada em fundamentos econômicos e financeiros.

3 METODOLOGIA

Esta seção apresenta a metodologia de execução do presente trabalho. A Seção 3.1 define o escopo e os objetivos deste trabalho, delineando as principais áreas de investigação. Em seguida, a Seção 3.2 descreve as abordagens e técnicas utilizadas para planejar e conduzir o desenvolvimento da ferramenta proposta. Por fim, a Seção 3.3 detalha os passos do trabalho, incluindo a investigação da estrutura dos dados, a modelagem lógica e o desenvolvimento da ferramenta de integração de dados.

3.1 Classificação da Pesquisa

Para fundamentar a classificação metodológica da pesquisa, foram consultadas as obras de **cervo:1983:metodologia**, que apresentam diferentes abordagens científicas, e de **gerhardt:2009:metodos**, que tratam de métodos amplamente utilizados na pesquisa acadêmica.

Quanto à natureza, esta pesquisa é aplicada, pois busca resolver um problema prático relacionado à integração automatizada de dados financeiros públicos disponibilizados pela CVM, promovendo avanços na análise fundamentalista e no apoio à tomada de decisão.

Em relação aos objetivos, a investigação é exploratória e descritiva. A parte exploratória está ligada à compreensão da estrutura dos dados da CVM, enquanto o aspecto descritivo se manifesta na proposição e implementação de uma ferramenta que permita o uso prático dessas informações de forma organizada.

Os procedimentos técnicos combinam pesquisa bibliográfica e documental para levantamento e compreensão teórica dos dados públicos, juntamente com o desenvolvimento de uma solução tecnológica que automatiza sua coleta, integração e disponibilização.

A abordagem metodológica adotada é mista. Aspectos qualitativos estão presentes na análise estrutural e na modelagem dos dados, enquanto os elementos quantitativos se refletem no processamento automatizado e na organização das informações em larga escala.

No contexto da área de Computação, a pesquisa representa o desenvolvimento de uma solução tecnológica voltada à resolução de um problema específico, com potencial de aplicação prática no meio acadêmico e no ambiente de análise financeira (**wazlawick:2009:metodo**).

3.2 Gerenciamento do Projeto

O gerenciamento deste projeto foi conduzido com base em práticas ágeis de desenvolvimento de software, adotando a metodologia *Scrum* como estrutura principal (**schwaber:2004:agile**). Como apoio visual, utilizou-se também o quadro de sinalização *Kanban*, contribuindo para a organização e o acompanhamento do fluxo de trabalho (**anderson:2010:kanban**).

Ambas as abordagens foram aplicadas de forma adaptada às necessidades do projeto, considerando sua natureza acadêmica e o escopo restrito da equipe envolvida. O *Scrum* foi utilizado de maneira personalizada, com ciclos iterativos de duas semanas, denominados *sprints*, nos quais as tarefas foram definidas, priorizadas e acompanhadas ao longo do desenvolvimento (**sutherland:2014:scrum**).

O Kanban complementou esse processo, funcionando como ferramenta visual para monitoramento das atividades, identificação de gargalos e ajustes de fluxo, contribuindo para uma gestão mais eficiente (**hammarberg:2014:kanban**). Essa combinação flexível e adaptada de métodos ágeis mostrou-se eficaz para garantir a organização, a visibilidade do progresso e a entrega incremental da solução proposta.

3.3 Solução proposta

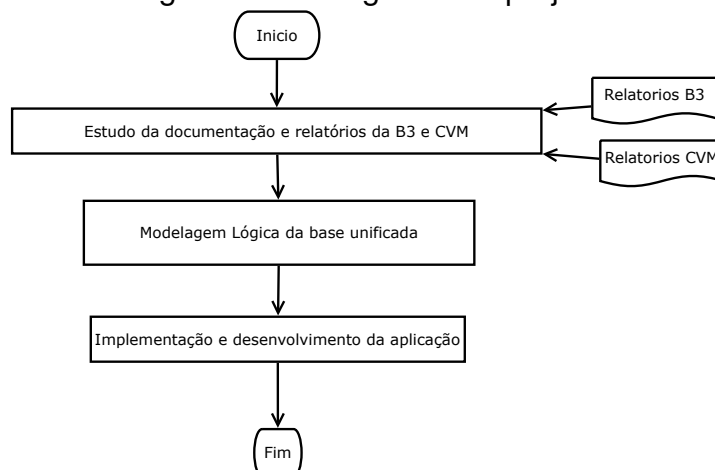
O ponto de partida do projeto consistiu em uma investigação detalhada da estrutura dos dados disponibilizados pela CVM. Essa etapa contemplou a análise dos formatos de arquivos, dos tipos de dados disponíveis e dos métodos de acesso às informações públicas fornecidas por essa instituição.

Embora os dados da B3 não tenham sido utilizados diretamente na construção da ferramenta, sua presença no contexto do mercado de capitais justificou a compreensão de sua estrutura e funcionamento. Dessa forma, foram consultados documentos como o relatório anual da CVM (**cvm:2023:relatorioanual**) e referências sobre a estrutura de dados da B3 (**b3:2023:estruturadados**), a fim de mapear desafios e possibilidades associados à integração de dados financeiros.

A Figura 2 ilustra o fluxo geral do projeto. A fase inicial compreendeu a coleta e documentação dos dados da CVM, seguida pela modelagem lógica que estruturou um banco de dados unificado, preparado para integração e análise posterior.

Com base na análise inicial, foi realizada a modelagem lógica da base de dados, com foco na criação de um esquema unificado que contemplasse as diferentes estruturas contábeis e financeiras presentes nos arquivos da CVM. Para essa etapa, foram consideradas boas práticas descritas na literatura sobre modelagem de dados financeiros (**domingues:2020:modelagem**) e integração de grandes volumes de dados (**perlin:2021:analise**).

Figura 2 – Fluxograma do projeto



Fonte: Elaborado pelo autor, 2024.

A etapa seguinte consistiu no desenvolvimento da aplicação de software responsável por automatizar o processo de coleta, tratamento e inserção dos dados na base unificada. Essa ferramenta foi desenvolvida com tecnologias modernas de extração e processamento de dados, garantindo flexibilidade e desempenho.

A aplicação foi capaz de realizar o download dos arquivos disponibilizados pela CVM, aplicar os tratamentos necessários para padronização e limpeza dos dados, e armazená-los de forma estruturada. Esse processo assegurou a integridade das informações e facilitou seu uso em análises fundamentalistas, estudos acadêmicos e outras aplicações. A proposta esteve alinhada com práticas consolidadas no desenvolvimento de soluções para automação e análise de dados (**prikladnicki:2014:desenvolvimentossoftware**) (**nhimi:2016:desenvolvimentossoftware**).

3.4 Materiais e Tecnologias

Esta seção apresenta os materiais e tecnologias que serão empregados durante o desenvolvimento do trabalho. O projeto será desenvolvido em um computador desktop cujas especificações são apresentadas no Quadro 1.

Durante a modelagem do banco de dados, utilizou-se o MySQL Workbench¹, versão 8.0. Essa ferramenta permitiu a criação visual do esquema lógico, facilitando o planejamento e a organização das entidades e relacionamentos.

Contudo, para o desenvolvimento da aplicação, optou-se pelo uso do SQLite², versão 3.49.1, como solução de banco de dados local. O SQLite foi adotado como banco de dados padrão do sistema, por oferecer maior praticidade na integra-

¹ <https://www.mysql.com/products/workbench/>

² <https://www.sqlite.org/>

Quadro 1 – Especificações do computador utilizado no TCC

| Componente | Especificação |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------|
| Processador | AMD Ryzen 5 5600G, 6-Core, 12-Threads, 3.6GHz (4.6GHz Turbo), Cache 19MB, AM4 |
| Memória RAM | 2x Kingston Fury Beast, 16GB, 3200MHz, DDR4, CL16 |
| Armazenamento | SSD TGT Seal ST, 240GB, SATA III, Leitura 500 MB/s, Gravação 450 MB/s HD WD Blue, 1TB, 3.5, 5400 RPM, SATA III, Cache 64MB |
| Sistema Operacional | Windows 11 Pro |

Fonte: Elaborado pelo autor, 2024.

ção com o código desenvolvido. Sua leveza, portabilidade e independência de um servidor o tornam especialmente adequado para o uso do código por estudantes e profissionais da área financeira.

A linguagem de programação utilizada foi o Python, na versão 3.12.9³, pela sua versatilidade, ampla comunidade e bibliotecas especializadas.

As bibliotecas utilizadas no projeto foram:

- Pandas, versão 2.2.1⁴
- Numpy, versão 1.26.4⁵
- BeautifulSoup4, versão 4.12.2⁶
- Lxml, versão 5.1.0⁷
- Requests, versão 2.31.0⁸
- SQLAlchemy, versão 2.0.27⁹
- Mysql-Connector-Python, versão 8.3.0¹⁰

A biblioteca *pandas* foi empregada para manipulação e análise de dados tabulares. Sua estrutura baseada em *DataFrames* proporcionou uma forma eficiente de organizar, filtrar e agrupar informações financeiras.

A biblioteca *numpy* foi utilizada em conjunto com o *pandas* para operações numéricas vetoriais e matriciais, fundamentais na realização de cálculos e no tratamento de grandes volumes de dados.

Para extração de informações de páginas web, foi utilizada a biblioteca *beautifulsoup4*, que permitiu navegar pela estrutura HTML de documentos e extrair dados de forma precisa. A ela foi integrada a biblioteca *lxml*, que serviu como parser rápido e eficiente de HTML e XML.

³ <https://www.python.org/downloads/release/python-3129/>

⁴ <https://pandas.pydata.org/>

⁵ <https://numpy.org/>

⁶ <https://www.crummy.com/software/BeautifulSoup/>

⁷ <https://lxml.de/>

⁸ <https://requests.readthedocs.io/en/latest/>

⁹ <https://www.sqlalchemy.org/>

¹⁰ <https://dev.mysql.com/downloads/connector/python/>

A coleta dos dados disponibilizados pela CVM foi realizada por meio da biblioteca *requests*, que viabilizou a comunicação com serviços HTTP e o download automatizado das informações necessárias.

A interface com o banco de dados foi desenvolvida com o auxílio do *SQLAlchemy*, que forneceu abstração e controle sobre as operações de inserção, consulta e atualização dos dados no SQLite, utilizando o paradigma de mapeamento objeto-relacional.

Durante a fase inicial de testes com a estrutura relacional modelada em MySQL, utilizou-se a biblioteca *mysql-connector-python*, conector oficial disponibilizado para integração entre aplicações Python e bancos de dados MySQL. Essa ferramenta assegurou plena compatibilidade entre o modelo relacional e os *scripts* desenvolvidos, permitindo validar o funcionamento da estrutura proposta em um ambiente robusto e amplamente utilizado no contexto de bancos relacionais.

Após essa etapa de validação, optou-se pela migração definitiva para o banco de dados SQLite. A escolha por essa tecnologia deve-se à sua leveza, portabilidade e facilidade de integração com sistemas locais, características particularmente vantajosas para estudos exploratórios, reproduzíveis e com foco acadêmico ou experimental, como é o caso deste trabalho.

4 DESENVOLVIMENTO

Este capítulo apresenta o processo de desenvolvimento da ferramenta proposta para a coleta, estruturação e integração de dados financeiros públicos disponibilizados pela CVM, com foco em análises fundamentalistas de companhias abertas brasileiras.

As atividades foram organizadas em três frentes principais, que estruturam as seções deste capítulo. A Seção 4.1 descreve a análise preliminar dos dados disponibilizados pela CVM, com destaque para a identificação de padrões, inconsistências e limitações nos metadados. A Seção 4.2 detalha a modelagem e estruturação da base de dados, com ênfase na padronização relacional e na aplicação de boas práticas de modelagem para contextos financeiros, conforme discutido na literatura (**elmasri:2016:fundamentals**). Por fim, a Seção 4.4 apresenta o sistema desenvolvido em Python, incluindo os módulos responsáveis pela coleta, transformação e armazenamento dos dados, bem como os mecanismos de registro de eventos (*logs*).

4.1 Análise inicial dos dados da CVM

A primeira etapa consistiu na compreensão do formato, da frequência de atualização e de outros aspectos relevantes relacionados aos dados disponibilizados pela CVM. Como o foco do trabalho são as Companhias (CIA) Abertas¹, realizou-se uma análise do site como fonte para obtenção dessas informações. A partir disso, foram identificados os conjuntos de dados referentes às Companhias Abertas, organizados nas seguintes pastas:

- informação cadastral;
- formulário cadastral;
- informações periódicas e eventuais;
- formulário de referência;
- valores mobiliários negociados e detidos;
- formulário de informações trimestrais;
- formulário de demonstrações financeiras padronizadas;
- informe do código de governança.

As informações cadastrais compõem uma categoria distinta, classificada como cadastro, enquanto os demais arquivos são agrupados sob a categoria de documentos. Ao acessar o site de dados da CVM², especificamente a seção referente às Companhias Abertas, é possível visualizar essas duas categorias de dados.

Será apresentada, a seguir, a descrição de cada pasta referente ao con-

¹ <https://dados.cvm.gov.br>

² <https://dados.cvm.gov.br/dados/>

junto de dados da CVM da CIA Aberta, com a finalidade de esclarecer o conteúdo e a utilidade de cada uma.

Adicionalmente, é importante contextualizar brevemente o sistema responsável pelo envio eletrônico dos documentos à CVM. O sistema *Empresas.NET* (ENET) é a plataforma oficial utilizada pelas companhias abertas para a transmissão de informações regulatórias exigidas pela autarquia. Essa ferramenta garante a padronização, segurança e rastreabilidade do envio, sendo o meio exclusivo de entrega de documentos como o formulário de referência, as demonstrações financeiras padronizadas, os informes trimestrais e demais obrigações periódicas previstas nas normas vigentes.

Os dados disponibilizados no portal da CVM são oferecidos em formatos estruturados em CSV (Comma-Separated Values), em texto simples no formato TXT e arquivos comprimidos no formato ZIP. Esses formatos visam garantir maior acessibilidade e automação no tratamento das informações. A maioria dos conjuntos de dados está organizada em arquivos anuais, separados por tipo de formulário e data de referência, o que permite consultas retroativas e reprodutibilidade dos resultados.

Essas pastas listadas anteriormente podem conter dados referentes diretamente à categoria principal ou subdivisões internas em outras palavras subcategorias, conforme a complexidade do conteúdo. Os metadados associados a cada categoria são apresentados de duas formas: quando não há subcategorias, os metadados são oferecidos em arquivos de texto; caso contrário, são fornecidos em arquivos compactados que, ao serem descompactados, revelam arquivos textual para cada subcategoria.

Os dados propriamente ditos também são disponibilizados em arquivos compactados por ano. Após a extração, o conteúdo pode variar: em categorias sem subdivisões, o arquivo anual tende a conter um único documento estruturado; já em categorias com subcategorias, o arquivo anual inclui múltiplos arquivos estruturado, organizados conforme os diferentes tipos de informação tratados.

Dessa forma, observa-se um padrão de organização hierárquica: pastas por categoria principal, contendo arquivos de metadados e dados, com granularidade ajustada conforme a existência de subcategorias. Essa estrutura visa facilitar o acesso modular à informação além de um controle sobre a atualização dos dados e permitir análises específicas com base no ano.

As legislações que fundamentam a disponibilização e a estrutura dos dados incluem, principalmente, a Resolução CVM nº 80/2022 (**cvm:2022:resolucao80**), que consolida as regras relativas ao registro e envio de informações pelas companhias abertas, e a Resolução CVM nº 44/2021 (**cvm:2021:resolucao44**), que dispõe sobre a divulgação de informações relativas a atos ou fatos relevantes, a negociação de valores mobiliários na pendência de tais informações ainda não divulgadas, bem como sobre a divulgação de operações realizadas com valores mobiliários por pessoas com acesso a informações privilegiadas. Tais normas conferem respaldo ju-

rídico à padronização e à publicidade dos dados disponibilizados, sendo essenciais para o desenvolvimento de sistemas automatizados de análise fundamentada, como o proposto neste trabalho.

4.1.1 Informação cadastral

O conjunto Informação Cadastral reúne os dados cadastrais das companhias abertas disponibilizados pela CVM. Entre as informações disponibilizadas, destacam-se o número do Cadastro Nacional da Pessoa Jurídica (CNPJ), a data de registro da companhia e a situação atual desse registro. Esses dados são fundamentais para análises regulatórias, econômicas e financeiras, além de servirem como base para estudos acadêmicos e pesquisas de mercado. O dicionário de dados, disponibilizado em formato textual, contém a descrição detalhada das colunas e dos tipos de dados presentes no arquivo principal. Já os dados propriamente ditos, que contêm os registros cadastrais das companhias abertas, são apresentados em um arquivo estruturado.

Este conjunto de dados é de acesso público e está disponível na plataforma de dados abertos da CVM. A base é atualizada diariamente.

Para facilitar a compreensão geral da estrutura e do conteúdo disponibilizado, apresenta-se no Apêndice A uma tabela-resumo com a visão geral dos campos e descrições desse conjunto de dados.

4.1.2 Formulário cadastral

O Formulário cadastral (FCA) é um documento eletrônico cuja entrega periódica ou eventual é regulamentada pela CVM e suas resoluções. Este formulário deve ser enviado à CVM por meio do sistema ENET, sendo uma obrigação regulatória destinada às companhias abertas. Sua principal função é atualizar e manter organizadas as informações institucionais e operacionais dessas entidades. O conjunto de dados públicos relacionados ao FCA, disponibilizado pela CVM, apresenta-se organizado em duas categorias principais de informações: os endereços de *download* dos documentos completos submetidos pelas companhias abertas e os conteúdos estruturados das seções que compõem o formulário.

Seguindo o padrão já apresentando de subcategorias o FCA possui a categoria refere-se ao arquivo `fca_cia_aberta`, que contém os links para o *download* dos documentos completos entregues pelas companhias nos últimos cinco anos. A visualização desses documentos requer a utilização do ENET. A as demais categoria compreende os conteúdos integrais do formulário, apresentados abaixo em arquivos estruturados por ano, estando dentro de um arquivo compactado por ano, conforme já

relatado anteriormente:

Figura 3 – Estrutura do FCA

```

/
├── fca_cia_aberta_geral
├── fca_cia_aberta_pais_estrangeiro_negociacao
├── fca_cia_aberta_canal_divulgacao
├── fca_cia_aberta_endereco
├── fca_cia_aberta_valor_mobiliario
├── fca_cia_aberta_auditor
├── fca_cia_aberta_escriturador
├── fca_cia_aberta_dri
└── fca_cia_aberta_departamento_acionistas

```

Fonte: Elaborado pelo autor, 2025.

Cada um desses arquivos aborda uma seção específica do formulário, possibilitando a segmentação e a análise detalhada das informações declaradas pelas companhias abertas. Uma visão geral dos arquivos estruturados do FCA, com a respectiva correspondência normativa e observações de uso, está apresentada no Apêndice B. Os arquivos são atualizados semanalmente, refletindo tanto novas entregas quanto reapresentações realizadas pelas companhias reguladas. Além disso, mantém-se um histórico contínuo desde o ano de 2010, incluindo inclusive arquivos que não estão sujeitos à política de atualização vigente. O dicionário de dados correspondente é disponibilizado em formato compactado e oferece descrições detalhadas de todas as colunas e variáveis presentes nos arquivos.

No desenvolvimento da ferramenta proposta neste trabalho, foi utilizado exclusivamente o arquivo `fca_cia_aberta_geral`, por concentrar de forma consolidada as principais informações institucionais das companhias abertas, o que dispensa a necessidade de integrar os demais arquivos individualizados.

Para fins de análise temporal e organização, os formulários referentes aos anos de 2020 a 2025 encontram-se agrupados em arquivos anuais distintos. Este conjunto de dados integra a base denominada *Documentos Periódicos e Eventuais de Regulados*. Sua manutenção ocorre de forma semanal, assegurando a atualização constante das informações disponibilizadas.

4.1.3 Informações periódicas e eventuais

O conjunto de dados referente às Informações Periódicas e Eventuais (IPE) reúne documentos não estruturados enviados por companhias abertas à CVM. Esses documentos, de caráter regulatório, contemplam tanto as obrigações periódicas quanto as comunicações eventuais exigidas ao longo das atividades societárias e da interlocução com o mercado. Trata-se de um acervo, que reflete a diversidade de

exigências legais e normativas aplicáveis às companhias reguladas. O conteúdo do conjunto está organizado em seis grandes categorias documentais:

- governança e estrutura societária;
- relação com investidores e mercado;
- informações econômico-financeiras e contábeis;
- transações e operações societárias;
- companhias em situação especial;
- informações regulatórias específicas.

Uma visão geral dos tipos documentais abrangidos pelo conjunto IPE pode ser consultada no Apêndice C, que apresenta uma tabela que descreve melhor essas categorias nos tipos de documentos presente no IPE.

A categoria de governança e estrutura societária compreende documentos que regulam a organização interna e as diretrizes de conduta das companhias, como Acordos de Acionistas, Estatutos Sociais, Regimentos Internos, Códigos de Conduta, além de Políticas de Governança e Sustentabilidade. Esses registros são fundamentais para garantir a conformidade, a transparência e a integridade das práticas de gestão corporativa.

Já os documentos de relação com investidores e mercado incluem comunicados oficiais direcionados ao público investidor e à sociedade, como Comunicados ao Mercado, Avisos a Acionistas e Debenturistas, Fatos Relevantes, Informações prestadas a bolsas estrangeiras e o Calendário de Eventos Corporativos. Tais documentos asseguram a divulgação tempestiva e adequada de informações relevantes, conforme exigido pelas normas da CVM.

No que se refere às informações econômico-financeiras e contábeis, o conjunto abrange dados sobre a situação patrimonial e os resultados das companhias, além de instrumentos financeiros e políticas de distribuição de lucros. Incluem-se nessa categoria Demonstrativos Financeiros, Documentos de Ofertas Públicas, Escrituras de Debêntures, Políticas de Dividendos e de Destinação de Resultados. A categoria de transações e operações societárias, por sua vez, contém registros de eventos relevantes que impactam a estrutura e o capital das empresas, como Comunicações sobre Transações com Partes Relacionadas, Ofertas Públicas de Aquisição (OPA) e Planos de Remuneração Baseados em Ações.

Já a documentação relativa a companhias em situação especial reúne informações sobre empresas que enfrentam processos de falência, liquidação ou recuperação judicial/extrajudicial, sendo essenciais para análise de risco e monitoramento da situação jurídico-financeira dessas entidades. Por fim, a categoria de informações regulatórias específicas abrange documentos exigidos por normativos da CVM, como as comunicações sobre negociação de valores mobiliários por pessoas ligadas, políticas de atuação de auditores independentes e outras políticas internas obrigatórias.

Todos os documentos do conjunto IPE são disponibilizados em arquivos compactados organizados por ano, mesmo quando cada compactado contém apenas um único arquivo estruturado em seu interior. Os arquivos correspondentes ao ano corrente e ao ano imediatamente anterior são atualizados semanalmente, incorporando tanto novas submissões quanto eventuais reapresentações. O acervo possui cobertura histórica desde 2003, incluindo arquivos legados que, embora não estejam sujeitos à política atual de atualização contínua, permanecem acessíveis para consulta.

Cada conjunto anual acompanha um arquivo auxiliar contendo o dicionário de dados em formato textual, no qual são descritas as variáveis e a lógica de indexação aplicadas à organização dos documentos. A relevância dos arquivos IPE transcende o aspecto meramente documental: sua análise permite identificar mudanças estruturais relevantes, como alterações no nome empresarial, na área de atuação ou na composição societária, constituindo uma fonte estratégica para o monitoramento de eventos corporativos vinculados a um determinado *ticker*.

4.1.4 Formulário de Referência

O Formulário de Referência (FRE) é um documento eletrônico cuja apresentação à CVM é obrigatória e periódica ou, em determinadas circunstâncias, eventual, por meio do sistema ENET. Sua principal finalidade é consolidar e divulgar, de maneira padronizada, um conjunto abrangente de informações sobre o emissor, contemplando sua estrutura societária, situação financeira, práticas de governança, riscos e relação com o mercado.

O conjunto de dados públicos associados ao FRE é composto por dois elementos principais:

- os endereços para *download* dos formulários submetidos;
- o conteúdo estruturado extraído desses documentos.

O primeiro item refere-se aos *links* diretos para os formulários completos entregues pelas companhias abertas nos últimos cinco anos, disponibilizados principalmente por meio do arquivo *fre_cia_aberta*. Já o segundo item corresponde aos dados tabulares derivados do conteúdo dos formulários, organizados em arquivos estruturados que abordam diversas dimensões informacionais, agrupadas nas seguintes categorias:

- informações institucionais e operacionais;
- aspectos financeiros e patrimoniais;
- administração e governança;
- ações e mercado;
- aspectos sociais e de diversidade.

Uma visão geral dos arquivos estruturados que compõem o conjunto FRE encontra-se apresentada no Apêndice D, com a descrição, referência normativa e indicação de uso no trabalho.

As informações estruturadas são atualizadas semanalmente, incorporando tanto novas submissões quanto reapresentações, e mantêm um histórico contínuo desde 2010. Essas atualizações derivam diretamente da base de Documentos Periódicos e Eventuais de Regulados submetidos à CVM. Cada pacote anual, disponível para os anos de 2010 a 2025, é disponibilizado em formato compactado e contém um conjunto completo de arquivos organizados por tipo de informação. Acompanha-se ainda um dicionário de dados, também compactado, que detalha as variáveis e colunas presentes.

Dentre os arquivos incluídos nesses pacotes, destaca-se o `fre_cia_aberta`, que possui especial relevância nesta pesquisa. Esse arquivo concentra, de forma consolidada, as principais informações estruturadas de cada submissão do FRE e inclui os identificadores e os links necessários para recuperação dos documentos completos no sistema ENET. Tal característica torna o `fre_cia_aberta` uma fonte centralizada e eficiente para acesso ao conteúdo essencial das companhias analisadas.

Embora os demais arquivos disponíveis nos pacotes — como `fre_cia_aberta_remuneracao_total_orgao`, `fre_cia_aberta_posicao_acionaria`, entre outros, ofereçam informações relevantes e detalhadas sobre aspectos específicos do formulário, sua estrutura fragmentada e granular, aliada à sobreposição parcial de informações com o arquivo principal, levou à opção metodológica por sua não integração nesta etapa do trabalho. Assim, adotou-se exclusivamente o `fre_cia_aberta` como base de dados para análise, considerando sua abrangência e consistência com os objetivos da pesquisa.

Essa delimitação visa garantir maior simplicidade no modelo de integração dos dados, sem prejuízo da possibilidade de incorporação futura de arquivos complementares, conforme a evolução das necessidades analíticas.

4.1.5 Valores mobiliários negociados e detidos

O conjunto de dados Valores mobiliários negociados e detidos (VLMO) contempla informações de envio obrigatório à CVM por meio do ENET. Trata-se de uma obrigação de natureza periódica, cujo cumprimento deve ser realizado pelas companhias abertas. Esse conjunto tem como objetivo registrar e divulgar, de maneira transparente, a posição e as negociações realizadas com valores mobiliários por administradores, membros do conselho fiscal, controladores e pessoas a eles vinculadas. O conjunto disponibiliza os informes entregues pelas companhias nos últimos

cinco anos, organizados em arquivos anuais compactados, contendo os dados estruturados. Esses arquivos reúnem informações como:

- nome e CPF/CNPJ dos declarantes;
- tipos e quantidades de valores mobiliários detidos;
- natureza da operação (compra, venda, bonificação, entre outros);
- data e características das transações realizadas;
- relação do declarante com a companhia emissora.

Os arquivos são acompanhados de um dicionário de dados, também em formato compactado. O conjunto é atualizado semanalmente. O histórico apesar de informado pela CVM que abrange os anos de 2020 a 2025, não está correto a informação pois é possível constatar dados de 2018. Ele utiliza como base os *Documentos Periódicos e Eventuais de Regulados*.

Importa esclarecer que, para este estudo, não foi utilizado nenhum arquivo do conjunto VLMO. Destaca-se, ainda, que esse conjunto assim como os demais analisados apresenta uma estrutura padronizada, composta por dois arquivos principais: um arquivo nomeado como *vlmo_cia_aberta*, que contém os endereços para *download* dos documentos entregues pelas companhias, e outro arquivo estruturado que apresenta a consolidação dos dados do VLMO. Ambos os arquivos encontram-se organizados por ano e disponibilizados em formato compactado.

4.1.6 Demonstrativos financeiros padronizados

Os demonstrativos financeiros padronizados (DFP) disponibilizados pela CVM, reúnem dados contábeis estruturados de companhias abertas brasileiras. Tanto o ITR quanto o DFP seguem um modelo uniforme de reporte, que inclui as principais demonstrações financeiras, informações cadastrais, pareceres e arquivos para *download*.

o conjunto de dados contempla as principais demonstrações financeiras, apresentadas em formato estruturado:

o conjunto de dados contempla as principais demonstrações financeiras, apresentadas em formato estruturado:

- balanço patrimonial ativo (BPA);
- balanço patrimonial passivo (BPP);
- demonstrações dos fluxos de caixa - método direto (DFC-MD);
- demonstrações dos fluxos de caixa - método indireto (DFC-MI);
- demonstração das mutações do patrimônio líquido (DMPL);
- demonstração do resultado (DRE);
- demonstração do resultado abrangente (DRA);
- demonstração do valor adicionado (DVA).

A principal distinção entre os dois documentos está na periodicidade, o ITR é divulgado trimestralmente, enquanto o DFP é publicado anualmente.

4.1.6.1 Informações trimestrais

As informações trimestrais (ITR) é um documento eletrônico de entrega obrigatória pelas companhias abertas à CVM. Ele reúne informações contábeis elaboradas de forma trimestral, em conformidade com as normas contábeis vigentes, com o objetivo de garantir a transparência e o acompanhamento contínuo do desempenho das empresas. Além das demonstrações financeiras, o ITR inclui declarações de responsáveis, dados cadastrais, composição do capital social e *links* para *download* dos arquivos completos. Os registros históricos remontam a 2011, e os dados são acompanhados por um dicionário de variáveis disponível em arquivos compactados.

4.1.6.2 Formulário de demonstrações financeiras padronizadas

O Formulário de demonstrações financeiras padronizada (DFP) reúne as demonstrações financeiras anuais das companhias abertas, conforme as resoluções apresentadas a cima, e é submetido exclusivamente por meio do sistema ENET. Esse formulário constitui uma base essencial para a análise da saúde patrimonial e econômica das empresas listadas no mercado de capitais brasileiro. O conjunto de dados é atualizado semanalmente e abrange informações desde 2010. Além das demonstrações financeiras, inclui pareceres de auditoria, declarações de administradores, dados cadastrais, composição acionária e *links* para *download* dos formulários. As demonstrações apresentam tanto contas fixas quanto variáveis, o que permite análises comparáveis e também personalizadas. Um dicionário técnico em formato compactado detalha todas as variáveis e campos disponíveis.

4.1.7 Informe do Código de Governança

O Informe do Código de Governança é um dos instrumentos regulatórios exigidos das companhias abertas brasileiras, com o objetivo de divulgar, de forma estruturada, as práticas de governança corporativa adotadas. Embora seja referenciado no site institucional da CVM pela sigla ICBGC, nos conjuntos de dados disponibilizados na plataforma de dados abertos, o conteúdo correspondente aparece sob a sigla CGVN. Considerando essa divergência de nomenclatura, este trabalho adota a sigla CGVN, conforme empregada nos arquivos estruturados.

O envio do CGVN é de caráter obrigatório e periódico, conforme estipulado pelas resoluções já apresentadas. O processo de entrega deve ser realizado

exclusivamente por meio do sistema ENET, utilizado pelas companhias abertas para o cumprimento de suas obrigações junto à CVM.

Esse informe tem como propósito promover a transparência das práticas de governança corporativa, oferecendo à sociedade, aos investidores e aos demais agentes de mercado uma visão clara e comparável sobre como as companhias organizam sua estrutura decisória, implementam mecanismos de controle e estabelecem políticas institucionais. Também permite avaliar o grau de aderência aos princípios fundamentais de governança, como equidade, responsabilidade corporativa, prestação de contas e transparência.

Os principais tópicos abordados no CGVN são:

- estrutura de governança vigente na companhia;
- composição e funcionamento dos órgãos de administração e fiscalização;
- políticas corporativas implementadas;
- nível de aderência às práticas recomendadas pelo Código Brasileiro de Governança Corporativa;
- justificativas apresentadas para os casos em que as recomendações não são seguidas.

O conteúdo informado pelas companhias é padronizado e estruturado para possibilitar análise automatizada em larga escala. Os dados abrangem o período de 2018 a 2025 e são atualizados semanalmente, contemplando tanto novas submissões quanto reapresentações.

O conjunto de dados CGVN é disponibilizado anualmente em arquivos compactados e organizados da seguinte forma:

- `cgvm_cia_aberta`;
- `cgvn_cia_aberta_praticas`.

O primeiro arquivo contém os endereços para *download* dos documentos completos entregues pelas companhias por meio do sistema ENET. Já o segundo arquivo apresenta os dados estruturados de forma tabular, organizados por companhia e prática declarada, permitindo processamento e análise automatizada.

O arquivo `cgvn_cia_aberta_praticas` inclui colunas que indicam o CNPJ da companhia, a data de referência do informe, a versão do documento, o nome empresarial, os identificadores de documento e item, além dos campos descritivos com o capítulo do código, o princípio relacionado, a prática recomendada, a indicação de adoção e, quando aplicável, a justificativa para sua não adoção.

Cabe destacar que, para os fins deste trabalho, o conjunto de dados CGVN não foi incluído na análise prática, uma vez que o foco recai sobre conjuntos mais diretamente relacionados às demonstrações financeiras e ao cadastro das companhias.

4.1.8 Resumo dos dados iniciais da CVM

Com o intuito de compreender a estrutura e o comportamento dos dados disponibilizados pela CVM, foram desenvolvidos dois scripts em Python, descritos nos Apêndices E e F. O primeiro script realiza a varredura recursiva na pasta pública de dados abertos da CVM, efetuando o *download* de todos os arquivos disponíveis sejam eles estruturados, compactados ou em formato textual sem a aplicação de filtros. O segundo script executa a extração em lote dos arquivos compactados, permitindo não apenas o acesso ao conteúdo interno, mas também a mensuração do volume, da diversidade e da granularidade dos dados presentes em cada conjunto. Essa etapa foi essencial para a análise preliminar, fornecendo uma visão concreta da organização e complexidade das bases tratadas.

Com base na análise exploratória viabilizada pelos *scripts* desenvolvidos, foi possível identificar quais conjuntos de dados da CVM apresentam maior relevância, cobertura e estrutura adequada aos objetivos deste trabalho. A seleção final concentrou-se em bases que oferecem informações estruturadas, atualizadas e historicamente abrangentes, com potencial para análises fundamentalistas automatizadas. O Quadro 2 resume os principais aspectos dos conjuntos de dados escolhidos, os quais constituem o núcleo informacional utilizado ao longo deste estudo.

Quadro 2 – Comparativo entre conjuntos de dados da CVM

| Aspecto | ITR | DFP | FRE | FCA | IPE |
|----------------------|----------------------------------|----------------------------------|----------------------------------------------|-----------------------------------|-----------------------------------------------------|
| Tipo de conteúdo | Informações Trimestrais | Demonstrações Financeiras Anuais | Formulário de Referência | Cadastro de Companhia Aberta | Documentos Periódicos/Eventuais |
| Frequência | Trimestral | Anual | Periódico/Eventual | Periódico/Eventual | Periódico/Eventual |
| Formato | CSV em ZIP | CSV em ZIP | CSV em ZIP | CSV em ZIP | CSV em ZIP |
| Volume descompactado | 9,62 GB | 3,49 GB | 941 MB | 28,2 MB | 261 MB |
| Estrutura | Demonstrativos contábeis | Demonstrativos consolidados | Informações qualitativas diversas | Dados cadastrais padronizados | Documentos PDF + metadados |
| Atualização | Semanal | Semanal | Semanal | Semanal | Semanal (A e A-1) |
| Período histórico | Desde 2011 | Desde 2010 | Desde 2010 | Desde 2010 | Desde 2003 |
| Destaques | BPA, DRE, DFC, DMPL, DVA etc. | BPA, DRE, DFC, DMPL, DVA etc. | Capital social, remuneração, ESG, governança | CNPJ, endereço, auditor, canais | Assembleias, fatos relevantes, estatutos, políticas |
| Complexidade | Alta (parse hierárquico) | Média (consolidação) | Alta (texto livre + variedade de temas) | Baixa | Alta (texto livre, reapresentações) |
| Utilização | Análise de desempenho trimestral | Análise patrimonial e histórica | Análise estratégica e institucional | Base de entidades (Dim_Companhia) | Monitoramento de eventos corporativos |

Fonte: Elaborado pelo autor, 2025.

A análise exploratória dos arquivos baixados e extraídos forneceu insumos valiosos para a etapa seguinte: o mapeamento dos dados. Essa fase consistiu na identificação das tabelas e colunas de origem, na avaliação de sua frequência de atualização e na estruturação de um modelo preliminar de destino. As informações

coletadas por meio dos metadados auxiliaram na definição de domínios, tipos, tamanhos e descrições de campos, estabelecendo as bases técnicas para a modelagem relacional tratada na próxima seção.

4.2 Modelagem e Estruturação dos Dados

A modelagem da base de dados teve início com o mapeamento das principais categorias disponibilizadas pela CVM — FCA, FRE, IPE, DFP e ITR — considerando também, quando existentes, suas respectivas subcategorias. Essa etapa preliminar teve como foco a análise da estrutura dos arquivos e dos metadados associados, permitindo compreender a organização, a granularidade e os padrões recorrentes nos dados fornecidos. O detalhamento desse processo encontra-se no Apêndice G. O procedimento adotado seguiu uma lógica semelhante à de um processo ETL (*Extract, Transform, Load*), com ênfase na padronização, organização e enriquecimento das informações extraídas das demonstrações financeiras reportadas por companhias abertas à CVM.

Durante a etapa de extração, foram utilizados arquivos textual que possuem os metadados, presente em arquivos compactados. Esses pacotes são atualizados semanalmente e organizados por tipo de demonstração contábil, como BPA, BPP, DRE, DFC-MD e DFC-MI, DMPL, DRA, DVA e pareceres de auditoria. Cada arquivo trata exclusivamente de um tipo de demonstração, garantindo segmentação e clareza na origem dos dados.

A etapa de transformação consistiu na padronização dos nomes de colunas e na normalização das estruturas tabulares. Realizou-se o mapeamento entre os campos originais e seus equivalentes no modelo relacional, adotando nomenclatura clara no padrão `snake_case`. Exemplos incluem a substituição de `CD_CVM` por `codigo_cvm` e `DS_CONTA` por `descricao_conta`. Adicionalmente, foi realizado o enriquecimento temporal dos dados por meio da criação de três campos derivados a partir da data de referência (`DT_REFER`): `data_doc`, `mes_doc` e `ano_doc`. A tipagem dos campos foi definida com base em seus domínios originais, utilizando tipos como `varchar`, `decimal`, `date`, `smallint` e `char`, com destaque para colunas binárias com valores padronizados como “S” para sim ou “N” para não.

O carregamento dos dados estruturados ocorreu de forma centralizada, com a maioria das informações sendo direcionada à tabela `Dfp`, responsável por armazenar os demonstrativos financeiros. Casos específicos, como pareceres de auditoria, foram alocados na tabela `Dfp_Parecer`. Adicionalmente, foi criada a tabela auxiliar `grupo_dfp`, destinada a indicar se os dados referem-se a demonstrativos individuais ou consolidados. As atualizações seguem a mesma periodicidade dos arquivos-fonte, com carga semanal.

A estrutura resultante desse processo é padronizada e favorece a integração em repositórios analíticos, como *data warehouses*, além de viabilizar consultas temporais e comparações entre companhias. Para isso, foram adotadas estratégias específicas de modelagem, como:

- redução de campos e tabelas desnecessárias;
- definição clara e consistente dos relacionamentos;
- unificação dos demonstrativos contábeis em tabelas únicas;
- criação de índices específicos para acelerar consultas analíticas.

Com base em reuniões periódicas com o orientador, definiu-se que o foco da modelagem seria a otimização da estrutura de dados com vistas à análise fundamentalista, priorizando desempenho, integridade e reprodutibilidade. A primeira versão do esquema foi desenvolvida no MySQL Workbench, conforme ilustrado no Apêndice H. Essa versão inicial foi utilizada como base para testes de estruturação lógica e verificação do comportamento das tabelas, embora ainda apresentasse redundâncias e ausência de refinamentos.

Na sequência, foram realizados ajustes sucessivos no modelo, conforme apresentado no Apêndice I, com o objetivo de eliminar complexidades desnecessárias e tabelas pouco relevantes ao escopo do trabalho. A versão intermediária, descrita no Apêndice J, já incorporava a maior parte das estratégias descritas, refletindo uma estrutura mais coerente com a finalidade do projeto.

Por fim, a versão final do modelo lógico foi implementada em SQLite e encontra-se documentada no Apêndice K. Essa versão consolida as decisões de modelagem e incorpora tabelas auxiliares que contribuiram para a redução do tamanho da base e para o ganho de eficiência nas operações. Trata-se do modelo utilizado na ferramenta desenvolvida e adotado como base da integração de dados no sistema proposto.

4.3 Características e Benefícios Analíticos da Estrutura Final

A estrutura final, fruto de sucessivos refinamentos, possui as seguintes características principais:

- Estrutura simples e clara;
- Facilidade de integração com sistemas externos;
- Disponibilidade pública e acessível.

A base contém apenas os dados essenciais, com nomenclaturas padronizadas e organização relacional intuitiva. Sua modularidade facilita a integração com outras fontes e ferramentas de análise.

Destaca-se ainda por ser aberta e pública, podendo ser utilizada em pesquisas acadêmicas, ferramentas de apoio à decisão financeira, além de iniciativas de

transparência e educação financeira.

Entre os principais benefícios analíticos proporcionados pela base estão:

- Facilidade na extração de indicadores financeiros;
- Avaliação precisa do endividamento;
- Estimativas confiáveis de valor intrínseco;
- Comparações históricas e setoriais consistentes;
- Apoio robusto à tomada de decisões financeiras.

Esses aspectos permitem análises fundamentalistas detalhadas, especialmente por meio da aplicação de indicadores de rentabilidade, como lucro por ação (LPA) e preço sobre lucro (P/L), além dos indicadores de liquidez, endividamento e das estimativas consistentes do valor intrínseco das empresas.

Por fim, a padronização dos dados possibilita comparações históricas e setoriais rigorosas, fornecendo um suporte consistente e confiável para investidores, pesquisadores e analistas financeiros.

4.4 Software

5 CONCLUSÃO

REFERÊNCIAS

GLOSSÁRIO

LaTeX – Linguagem de marcação utilizada principalmente para a composição de documentos técnicos e científicos, fornecendo uma formatação consistente e de alta qualidade.

Modelo / Template – Documento ou conjunto de elementos predefinidos que serve de estrutura base para a criação de outros documentos, permitindo uma formatação consistente e facilitando o trabalho de edição.

APÊNDICES

APÊNDICE A - DICIONÁRIO RESUMIDO — INFORMAÇÃO CADASTRAL

A Tabela 1 apresenta uma visão geral dos campos disponíveis no conjunto informação cadastral, conforme publicado no portal de dados abertos da CVM. Para cada campo, são indicados a descrição, o tipo de dado e o respectivo tamanho ou precisão.

Tabela 1 – Resumo dos campos da base Informação Cadastral

| Campo | Descrição | Tipo de Dados | Tam./Prec. |
|---------------------|--------------------------------------|----------------------|-------------------|
| AUDITOR | Nome do Auditor | varchar | 100 |
| BAIRRO | Bairro | varchar | 100 |
| BAIRRO_RESP | Bairro do responsável | varchar | 100 |
| CATEG_REG | Categoria do registro | varchar | 20 |
| CD_CVM | Código CVM | numeric | 7 |
| CEP | CEP | numeric | 8 |
| CEP_RESP | CEP do responsável | numeric | 8 |
| CNPJ_AUDITOR | CNPJ do Auditor | varchar | 20 |
| CNPJ_CIA | CNPJ da companhia | varchar | 20 |
| COMPL | Complemento de endereço | varchar | 100 |
| COMPL_RESP | Complemento do responsável | varchar | 100 |
| CONTROLE_ACIIONARIO | Controle Acionário | varchar | 30 |
| DDD_FAX | Código DDD (FAX) | varchar | 4 |
| DDD_FAX_RESP | Código DDD (FAX) do responsável | varchar | 4 |
| DDD_TEL | Código DDD (Telefone) | varchar | 4 |
| DDD_TEL_RESP | Código DDD (Telefone) do responsável | varchar | 4 |
| DENOM_COMERC | Denominação Comercial | varchar | 100 |
| DENOM_SOCIAL | Denominação Social | varchar | 100 |
| DT_CANCEL | Data de cancelamento | date | 10 |
| DT_CONST | Data de constituição | date | 10 |

| Campo | Descrição | Tipo de Dados | Tam./Prec. |
|--------------------|---------------------------------|----------------------|-------------------|
| DT_INI_CATEG | Início da categoria do registro | date | 10 |
| DT_INI_RESP | Início do responsável | date | 10 |
| DT_INI_SIT | Início da situação | date | 10 |
| DT_INI_SIT_EMISSOR | Início da situação do emissor | date | 10 |
| DT_REG | Data de registro | date | 10 |
| EMAIL | E-mail | varchar | 100 |
| EMAIL_RESP | E-mail do responsável | varchar | 100 |
| FAX | FAX | numeric | 15 |
| FAX_RESP | FAX do responsável | numeric | 15 |
| LOGRADOURO | Logradouro | varchar | 100 |
| LOGRADOURO_RESP | Logradouro do responsável | varchar | 100 |
| MOTIVO_CANCEL | Motivo de cancelamento | varchar | 80 |
| MUN | Município | varchar | 100 |
| MUN_RESP | Município do responsável | varchar | 100 |
| PAIS | País | varchar | 100 |
| PAIS_RESP | País do responsável | varchar | 100 |
| RESP | Nome do responsável | varchar | 100 |
| SETOR_ATIV | Setor de atividade | varchar | 100 |
| SIT | Situação | varchar | 40 |
| SIT_EMISSOR | Situação do emissor | char | 80 |
| TEL | Telefone | numeric | 15 |
| TEL_RESP | Telefone do responsável | numeric | 15 |
| TP_ENDER | Tipo de endereço | char | 30 |
| TP_MERC | Tipo de mercado | varchar | 50 |
| TP_RESP | Tipo de responsável | varchar | 100 |
| UF | Unidade da Federação | char | 2 |
| UF_RESP | UF do responsável | char | 2 |

Fonte: Elaborado pelo autor, 2024.

APÊNDICE B - RESUMO DO FORMULÁRIO CADASTRAL

A Tabela 2 apresenta um panorama consolidado dos arquivos estruturados que compõem o Formulário Cadastral (FCA), indicando suas respectivas seções no Anexo 22 da Instrução CVM nº 480, a descrição das informações contidas e eventuais observações relevantes para este trabalho.

Tabela 2 – Arquivos Estruturados do Formulário Cadastral (FCA)

| Arquivo | Descrição | Seção / Item | Uso no Trabalho |
|--------------------------------------------|-----------------------------------------------------------------------|-------------------|-----------------|
| fca_cia_aberta_geral | Informações institucionais e gerais da companhia | Seção 1 | Utilizado |
| fca_cia_aberta_pais_estrangeiro_negociacao | Países onde a companhia possui negociação de valores mobiliários | Itens 1.14 e 1.15 | Não |
| fca_cia_aberta_canal_divulgacao | Canais utilizados para divulgação de informações ao mercado | Item 1.24 | Não |
| fca_cia_aberta_endereco | Endereços físicos e contatos institucionais | Itens 1.25 a 1.30 | Não |
| fca_cia_aberta_valor_mobiliario | Detalhamento dos valores mobiliários emitidos pela companhia | Seção 2 | Não |
| fca_cia_aberta_auditor | Informações sobre o auditor independente contratado | Seção 3 | Não |
| fca_cia_aberta_escriturador | Dados sobre o escriturador de ações da companhia | Seção 4 | Não |
| fca_cia_aberta_dri | Informações sobre o responsável pelas relações com investidores (DRI) | Seção 5 | Não |

| Arquivo | Descrição | Seção / Item | Uso no Trabalho |
|--------------------------------------------|-----------------------------------------------------------------|--------------|-----------------|
| fca_cia_aberta_ departamento_acionistas | Estrutura e dados do departamento de atendimento aos acionistas | Seção 6 | Não |

Fonte: Elaborado pelo autor, 2024.

APÊNDICE C - LISTA DE DOCUMENTOS DISPONÍVEIS NO CONJUNTO IPE

Tabela 3 – Documentos Estruturados das Informações Periódicas e Eventuais (IPE)

| Documento | Descrição | Categoria Geral |
|-------------------------------------------------------|-----------------------------------------------------|------------------------|
| Acordo de Acionistas | Regras pactuadas entre sócios sobre controle e voto | Governança |
| Assembleia | Atas e deliberações de assembleias gerais | Governança |
| Aviso aos Acionistas | Informações operacionais e societárias relevantes | Investidores |
| Aviso aos Debenturistas | Comunicados voltados aos detentores de debêntures | Investidores |
| Calendário de Eventos Corporativos | Agenda de eventos obrigatórios da companhia | Investidores |
| Carta Anual de Governança Corporativa | Declaração sobre práticas de governança | Governança |
| Código de Conduta | Normas internas de comportamento e ética | Governança |
| Comunicado ao Mercado | Divulgações públicas não obrigatórias | Investidores |
| Fato Relevante | Divulgação obrigatória de eventos materiais | Investidores |
| Informações sobre Falência, Liquidação ou Recuperação | Status jurídico-financeiro da companhia | Situação Especial |
| Plano de Remuneração Baseado em Ações | Programas de incentivos aos executivos | Transações |
| Política de Divulgação | Normas sobre divulgação de informações sensíveis | Regulatória |
| Política de Sustentabilidade | Diretrizes para práticas sustentáveis | Governança |
| Regimento Interno do Conselho de Administração | Regras de funcionamento do conselho | Governança |

| Documento | Descrição | Categoria Geral |
|-----------------------------------------|----------------------------------------------------|------------------------|
| Edital de Oferta Pública de Ações (OPA) | Oferta para aquisição de ações no mercado | Transações |
| Escritura de Debêntures | Termos das emissões de debêntures | Econômico-financeira |
| Documentos de Oferta Pública | Informações para investidores em IPOs e follow-ons | Econômico-financeira |
| Informações às Bolsas Estrangeiras | Relatórios encaminhados a bolsas internacionais | Investidores |
| Política de Dividendos | Critérios para distribuição de lucros | Econômico-financeira |
| Outras Políticas Regulatórias | Conjunto de normas exigidas pela CVM | Regulatória |

Fonte: Elaborado pelo autor, 2024.

APÊNDICE D - ARQUIVOS ESTRUTURADOS DO FRE

Tabela 4 – Arquivos Estruturados do FRE

| Arquivo | Descrição | Seção / Item | Uso no Trabalho |
|----------------------------------------------------|-------------------------------------------------------------------|---------------------|------------------------|
| fre_cia_aberta | Metadados principais e links para download dos documentos no ENET | Geral | Utilizado |
| fre_cia_responsavel | Identificação dos responsáveis pelo formulário | Item 1.1 | Não |
| fre_cia_auditor | Informações sobre os auditores independentes | Itens 2.1/2 | Não |
| fre_cia_auditor_responsavel | Identificação do responsável técnico do auditor | Itens 2.1/2 | Não |
| fre_cia_aberta_informacao_financeira | Indicadores e dados financeiros consolidados | Item 3.1 | Não |
| fre_cia_aberta_distribuicao_dividendos | Política e histórico de distribuição de dividendos | Item 3.5 | Não |
| fre_cia_aberta_distribuicao_dividendos_classe_acao | Dividendos por classe de ação | Item 3.5 | Não |
| fre_cia_aberta_endividamento | Indicadores de endividamento da companhia | Item 3.7 | Não |
| fre_cia_aberta_obrigacao | Obrigações e dívidas em aberto | Item 3.8 | Não |
| fre_cia_historico_emissor | Constituição e registros históricos do emissor | Itens 6.1/2/4 | Não |
| fre_cia_grupo_economico_reestruturacao | Reestruturações societárias (descontinuado após 2016) | Item 8.3 | Não |
| fre_cia_ativo_imobilizado | Detalhes dos ativos imobilizados | Item 9.1.a | Não |
| fre_cia_ativo_intangivel | Detalhes dos ativos intangíveis | Item 9.1.b | Não |

| Arquivo | Descrição | Seção / Item | Uso no Trabalho |
|-----------------------------------------------------|--------------------------------------------------|---------------------|------------------------|
| fre_cia_participacao_sociedade | Participações societárias do emissor | Item 9.1.c | Não |
| fre_cia_participacao_sociedade_valorizacao_acao | Histórico de valorização de participações | Item 9.1.c | Não |
| fre_cia_aberta_administrador_membro_conselho_fiscal | Composição dos administradores e conselho fiscal | Itens 12.5/6 | Não |
| fre_cia_aberta_membro_comite | Composição dos comitês | Itens 12.7/8 | Não |
| fre_cia_aberta_relacao_familiar | Relações familiares entre administradores | Item 12.9 | Não |
| fre_cia_aberta_relacao_subordinacao | Relações de subordinação entre administradores | Item 12.10 | Não |
| fre_cia_aberta_remuneracao_total_orgao | Remuneração total dos órgãos administrativos | Item 13.2 | Não |
| fre_cia_aberta_remuneracao_maxima_minima_media | Remuneração máxima, mínima e média | Item 13.11 | Não |
| fre_cia_aberta_posicao_acionaria | Composição acionária consolidada | Item 15.1/2 | Não |
| fre_cia_aberta_posicao_acionaria_classe_acao | Composição acionária por classe de ação | Item 15.1/2 | Não |
| fre_cia_aberta_distribuicao_capital | Estrutura do capital social | Item 15.3 | Não |
| fre_cia_aberta_distribuicao_capital_classe_acao | Capital por classe de ação | Item 15.3.d | Não |
| fre_cia_aberta_transacao_parte_relacionada | Transações com partes relacionadas | Item 16.2 | Não |
| fre_cia_capital_social | Capital social emitido | Item 17.1 | Não |

| Arquivo | Descrição | Seção / Item | Uso no Trabalho |
|--------------------------------------------------|------------------------------------------|---------------------|------------------------|
| fre_cia_capital_social_classe_acao | Capital por classe de ação | Item 17.1 | Não |
| fre_cia_capital_social_titulo_conversivel | Títulos conversíveis em ações | Item 17.1 | Não |
| fre_cia_capital_social_aumento | Aumentos de capital social | Item 17.2 | Não |
| fre_cia_capital_social_aumento_classe_acao | Aumentos por classe de ação | Item 17.2 | Não |
| fre_cia_capital_social_desdobramento | Desdobramento, grupo-mento e bonificação | Item 17.3 | Não |
| fre_cia_capital_social_desdobramento_classe_acao | Desdobramento por classe de ação | Item 17.3 | Não |
| fre_cia_capital_social_reducao | Redução do capital social | Item 17.4 | Não |
| fre_cia_capital_social_reducao_classe_acao | Redução por classe de ação | Item 17.4 | Não |
| fre_cia_direito_acao | Direitos das ações | Item 18.1 | Não |
| fre_cia_volume_valor_mobiliario | Volume e cotação de valores mobiliários | Item 18.4 | Não |

Fonte: Elaborado pelo autor, 2024.

APÊNDICE E - CÓDIGO PARA BAIXAR OS DADOS DA CVM

```

import requests
from bs4 import BeautifulSoup
import os
import time

# Função para fazer o download do arquivo
def download_file(url, save_path):
    response = requests.get(url)
    with open(save_path, 'wb') as f:
        f.write(response.content)

# Função para fazer uma requisição HTTP com retries
def make_request(url, retries=5, delay=5):
    for i in range(retries):
        try:
            response = requests.get(url)
            response.raise_for_status()
            return response
        except requests.exceptions.RequestException as e:
            print(f"Erro ao acessar {url}: {e}")
            if i < retries - 1:
                print(f"Tentando novamente em {delay} segundos...")
                time.sleep(delay)
            else:
                raise

# URL base inicial
base_url = 'https://dados.cvm.gov.br/dados/CIA_ABERTA/'

# Diretório base de destino
dest_dir_base = 'CIA_ABERTA'

# Lista para armazenar os diretórios a serem explorados, iniciando com o URL base
dirs_to_explore = [base_url]

# Loop para explorar os diretórios
while dirs_to_explore:
    # Remove o primeiro diretório da lista
    current_dir = dirs_to_explore.pop(0)

    # Faz a requisição HTTP para obter o conteúdo da página
    response = make_request(current_dir)
    soup = BeautifulSoup(response.text, 'html.parser')

    # Encontra todos os links na página
    for link in soup.find_all('a'):
        href = link.get('href')

        # Ignora os links que são de volta ('../')
        if href == '../':
            continue

        # Constrói a URL completa do link
        full_url = current_dir + href

        # Se o link for um diretório, adiciona à lista de diretórios a serem explorados
        if href.endswith('/'):
            dirs_to_explore.append(full_url)
        else:
            # Se o link for um arquivo, verifica se é um arquivo CSV, ZIP ou TXT e faz o download
            if href.endswith('.csv') or href.endswith('.zip') or href.endswith('.txt'):
                # Obtém o caminho relativo do arquivo em relação à URL base
                relative_path = os.path.relpath(full_url, base_url)
                # Constrói o caminho completo de destino
                save_dir = os.path.join(dest_dir_base, os.path.dirname(relative_path))
                # Cria os diretórios necessários se não existirem
                os.makedirs(save_dir, exist_ok=True)
                # Constrói o caminho completo do arquivo de destino
                save_path = os.path.join(save_dir, os.path.basename(full_url))
                # Faz o download do arquivo
                download_file(full_url, save_path)
                print("Arquivo baixado:", save_path)

    # Imprime o diretório atual
    print("Diretório atual:", current_dir)

```


APÊNDICE F - CÓDIGO PARA EXTRAIR OS DADOS DA CVM

```
1 import os
2 import zipfile
3
4 # Caminhos de origem e destino
5 pasta_zip = r"CIA_ABERTA"
6 pasta_destino = os.path.join(os.path.dirname(pasta_zip), "CIA_ABERTA_extraida")
7
8 # Percorrer a pasta de origem
9 for root, _, files in os.walk(pasta_zip):
10     for file in files:
11         if file.endswith(".zip"):
12             caminho_zip = os.path.join(root, file)
13
14             # Caminho relativo da subpasta em relação à raiz da pasta_zip
15             caminho_relativo = os.path.relpath(root, pasta_zip)
16
17             # Criar o caminho correspondente na pasta de destino
18             destino_subpasta = os.path.join(pasta_destino, caminho_relativo)
19             os.makedirs(destino_subpasta, exist_ok=True)
20
21             # Extrair o conteúdo do ZIP nesse caminho
22             try:
23                 with zipfile.ZipFile(caminho_zip, 'r') as zip_ref:
24                     zip_ref.extractall(destino_subpasta)
25                     print(f"Extraído: {file} para {destino_subpasta}")
26             except zipfile.BadZipFile:
27                 print(f"Erro: Arquivo ZIP inválido - {caminho_zip}")
```

APÊNDICE G - MAPEAMENTO COMPLETO DOS DADOS DA CVM

| | | | | | | |
|---------------------------------|------------------------------|-----------------------------------------------------------------------------------|---------------|--------------|------------------|-------------------|
| DFP - Dados Cadastrais | | | | | | |
| Arquivo de Origem | Periodicidade de atualização | Tabela de Destino | | | | |
| meta_dfp_cia_aberta_txt.zip | Semanal | Dfp | | | | |
| meta_dfp_cia_aberta.txt | | | | | | |
| | | | | | | |
| Colunas Origem | Colunas Destino | Descrição coluna | Tipo de dados | Domínio | Tamanho/Precisão | |
| CATEG_DOC | categoria_doc | Categoria do documento | varchar | Alfanumérico | 20 | |
| CD_CVM | codigo_cvm | Código CVM | char | Numérico | 6 | |
| CNPJ_CIA | cnnpj_companhia | CNPJ da companhia | varchar | Alfanumérico | 20 | |
| DENOM_CIA | denominacao_companhia | Nome empresarial da companhia | varchar | Alfanumérico | 100 | |
| DT_RECEB | data_recebimento_doc | Data da recebimento do documento | date | AAAA-MM-DD | 10 | |
| DT_REFER | data_referencia_doc | Data de referência do documento | date | AAAA-MM-DD | 10 | |
| ID_DOC | id_doc | Identificador do documento | int | Numérico | 10 | |
| LINK_DOC | link_doc | Endereço para download do documento | varchar | Alfanumérico | 121 | |
| VERSAO | versao | Versão do documento | smallint | Numérico | 5 | |
| | data_doc | Data do documento (aspecto temporal) | date | AAAA-MM-DD | 10 | |
| | mes_doc | Mês do documento (aspecto temporal) | varchar | Numérico | 4 | |
| | ano_doc | Ano do documento (aspecto temporal) | varchar | Numérico | 4 | |
| | | | | | | |
| | | | | | | |
| Balanco Patrimonial Ativo (BPA) | | | | | | |
| Arquivo de Origem | Periodicidade de atualização | | | | | |
| meta_dfp_cia_aberta_txt.zip | Semanal | | | | | |
| meta_dfp_cia_aberta_BPA.txt | | | | | | |
| | | | | | | |
| Colunas Origem | Colunas Destino | Descrição coluna | Tipo de dados | Domínio | Tamanho/Precisão | Tabela de Destino |
| CD_CONTA | codigo_conta | Código da conta | varchar | Numérico | 18 | Dfp |
| CD_CVM | codigo_cvm | Código CVM | char | Numérico | 6 | Dfp |
| CNPJ_CIA | cnnpj_companhia | CNPJ da companhia | varchar | Alfanumérico | 20 | |
| DENOM_CIA | denominacao_companhia | Nome empresarial da companhia | varchar | Alfanumérico | 100 | |
| DS_CONTA | descricao_conta | Descrição da conta | varchar | Alfanumérico | 100 | |
| DT_FIM_EXERC | data_fim_exercicio | Data fim do exercicio social | date | AAAA-MM-DD | 10 | Dfp |
| DT_REFER | data_referencia_doc | Data de referência do documento | date | AAAA-MM-DD | 10 | Dfp |
| ESCALA_MOEDA | escala_monetaria | Escala monetária | varchar | Alfanumérico | 100 | Dfp |
| GRUPO_DFP | grupo_dfp | Nome e nível de agregação da demonstração (Mostra se é Individual ou Consolidado) | varchar | Alfanumérico | 206 | grupo_dfp |
| MOEDA | moeda | Moeda | varchar | Alfanumérico | 100 | Dfp |
| ORDEM_EXERC | ordem_exercicio | Ordem do exercicio social | varchar | Alfanumérico | 9 | Dfp |
| ST_CONTA_FIXA | tipo_conta | Indica se é conta fixa ou não | varchar | S/N | 1 | Dfp |
| VERSAO | versao | Versão do documento | smallint | Numérico | 5 | Dfp |
| VL_CONTA | valor_conta | Valor da conta | decimal | Numérico | 29 | Dfp |
| | data_doc | Data do documento (aspecto temporal) | date | AAAA-MM-DD | 10 | Dfp |
| | mes_doc | Mês do documento (aspecto temporal) | varchar | Numérico | 4 | Dfp |
| | ano_doc | Ano do documento (aspecto temporal) | varchar | Numérico | 4 | Dfp |
| | | | | | | |
| | | | | | | |
| Balanco Patrimonial Ativo (BPP) | | | | | | |
| Arquivo de Origem | Periodicidade de atualização | | | | | |
| meta_dfp_cia_aberta_txt.zip | Semanal | | | | | |
| meta_dfp_cia_aberta_BPP.txt | | | | | | |
| | | | | | | |

| Colunas Origem | Colunas Destino | Descrição coluna | Tipo de dados | Domínio | Tamanho/Precisão | Tabela de Destino |
|-----------------------------------------------------------|-----------------------|-----------------------------------------------------------------------------------|---------------|--------------|------------------|-------------------|
| CD_CONTA | codigo_conta | Código da conta | varchar | Númérico | 18 | Dfp |
| CD_CVM | codigo_cvm | Código CVM | char | Númérico | 6 | Dfp |
| CNPJ_CIA | cnpj_companhia | CNPJ da companhia | varchar | Alfanumérico | 20 | |
| DENOM_CIA | denominacao_companhia | Nome empresarial da companhia | varchar | Alfanumérico | 100 | |
| DS_CONTA | descricao_cotna | Descrição da conta | varchar | Alfanumérico | 100 | |
| DT_FIM_EXERC | data_fim_exercicio | Data fim do exercicio social | date | AAAA-MM-DD | 10 | Dfp |
| DT_REFER | data_referencia_doc | Data de referência do documento | date | AAAA-MM-DD | 10 | Dfp |
| ESCALA_MOEDA | escala_monetaria | Escala monetária | varchar | Alfanumérico | 100 | Dfp |
| GRUPO_DFP | grupo_dfp | Nome e nível de agregação da demonstração (Mostra se é Individual ou Consolidado) | varchar | Alfanumérico | 206 | grupo_dfp |
| MOEDA | moeda | Moeda | varchar | Alfanumérico | 100 | Dfp |
| ORDEM_EXERC | ordem_exercicio | Ordem do exercicio social | varchar | Alfanumérico | 9 | Dfp |
| ST_CONTA_FIXA | conta_fixa | Indica se é conta fixa ou não | varchar | S/N | 1 | Dfp |
| VERSAO | versao | Versão do documento | smallint | Númérico | 5 | Dfp |
| VL_CONTA | valor_conta | Valor da conta | decimal | Númérico | 29 | Dfp |
| | data_doc | Data do documento (aspecto temporal) | date | AAAA-MM-DD | 10 | Dfp |
| | mes_doc | Mês do documento (aspecto temporal) | varchar | Númérico | 4 | Dfp |
| | ano_doc | Ano do documento (aspecto temporal) | varchar | Númérico | 4 | Dfp |
| | | | | | | |
| | | | | | | |
| Demonstração de Fluxo de Caixa - Método Direto (DFC-MD) | | | | | | |
| Arquivo de Origem | | Periodicidade de atualização | | | | |
| meta_dfp_cia_aberta_txt.zip | | Semanal | | | | |
| meta_dfp_cia_aberta_DFC_MD | | | | | | |
| | | | | | | |
| Colunas Origem | Colunas Destino | Descrição coluna | Tipo de dados | Domínio | Tamanho/Precisão | Tabela de Destino |
| CD_CONTA | codigo_conta | Código da conta | varchar | Númérico | 18 | Dfp |
| CD_CVM | codigo_cvm | Código CVM | char | Númérico | 6 | Dfp |
| CNPJ_CIA | cnpj_companhia | CNPJ da companhia | varchar | Alfanumérico | 20 | |
| DENOM_CIA | denominacao_companhia | Nome empresarial da companhia | varchar | Alfanumérico | 100 | |
| DS_CONTA | descricao_cotna | Descrição da conta | varchar | Alfanumérico | 100 | |
| DT_FIM_EXERC | data_fim_exercicio | Data fim do exercicio social | date | AAAA-MM-DD | 10 | Dfp |
| DT_INI_EXERC | data_inicio_exercicio | Data inicio do exercicio social | date | AAAA-MM-DD | 10 | Dfp |
| DT_REFER | data_referencia_doc | Data de referência do documento | date | AAAA-MM-DD | 10 | Dfp |
| ESCALA_MOEDA | escala_monetaria | Escala monetária | varchar | Alfanumérico | 100 | Dfp |
| GRUPO_DFP | grupo_dfp | Nome e nível de agregação da demonstração (Mostra se é Individual ou Consolidado) | varchar | Alfanumérico | 206 | grupo_dfp |
| MOEDA | moeda | Moeda | varchar | Alfanumérico | 100 | Dfp |
| ORDEM_EXERC | ordem_exercicio | Ordem do exercicio social | varchar | Alfanumérico | 9 | Dfp |
| ST_CONTA_FIXA | conta_fixa | Indica se é conta fixa ou não | varchar | S/N | 1 | Dfp |
| VERSAO | versao | Versão do documento | smallint | Númérico | 5 | Dfp |
| VL_CONTA | valor_conta | Valor da conta | decimal | Númérico | 29 | Dfp |
| | data_doc | Data do documento (aspecto temporal) | date | AAAA-MM-DD | 10 | Dfp |
| | mes_doc | Mês do documento (aspecto temporal) | varchar | Númérico | 4 | Dfp |
| | ano_doc | Ano do documento (aspecto temporal) | varchar | Númérico | 4 | Dfp |
| | | | | | | |
| | | | | | | |
| Demonstração de Fluxo de Caixa - Método Indireto (DFC-MI) | | | | | | |
| Arquivo de Origem | | Periodicidade de atualização | | | | |
| meta_dfp_cia_aberta_txt.zip | | Semanal | | | | |

meta_dfp_cia_aberta_DFC_MI.txt

| Colunas Origem | Colunas Destino | Descrição coluna | Tipo de dados | Dominio | Tamanho/Precisão | Tabela de Destino |
|----------------|-----------------------|-----------------------------------------------------------------------------------|---------------|--------------|------------------|-------------------|
| CD_CONTA | codigo_conta | Código da conta | varchar | Númerico | 18 | Dfp |
| CD_CVM | codigo_cvm | Código CVM | char | Númerico | 6 | Dfp |
| CNPJ_CIA | cnpj_companhia | CNPJ da companhia | varchar | Alfanumérico | 20 | |
| DENOM_CIA | denominacao_companhia | Nome empresarial da companhia | varchar | Alfanumérico | 100 | |
| DS_CONTA | descricao_cotna | Descrição da conta | varchar | Alfanumérico | 100 | |
| DT_FIM_EXERC | data_fim_exercicio | Data fim do exercicio social | date | AAAA-MM-DD | 10 | Dfp |
| DT_INI_EXERC | data_inicio_exercicio | Data inicio do exercicio social | date | AAAA-MM-DD | 10 | Dfp |
| DT_REFER | data_referencia_doc | Data de referência do documento | date | AAAA-MM-DD | 10 | Dfp |
| ESCALA_MOEDA | escala_monetaria | Escala monetária | varchar | Alfanumérico | 100 | Dfp |
| GRUPO_DFP | grupo_dfp | Nome e nível de agregação da demonstração (Mostra se é Individual ou Consolidado) | varchar | Alfanumérico | 206 | grupo_dfp |
| MOEDA | moeda | Moeda | varchar | Alfanumérico | 100 | Dfp |
| ORDEM_EXERC | ordem_exercicio | Ordem do exercicio social | varchar | Alfanumérico | 9 | Dfp |
| ST_CONTA_FIXA | conta_fixa | Indica se é conta fixa ou não | varchar | S/N | 1 | Dfp |
| VERSAO | versao | Versão do documento | smallint | Númerico | 5 | Dfp |
| VL_CONTA | valor_conta | Valor da conta | decimal | Númerico | 29 | Dfp |
| | data_doc | Data do documento (aspecto temporal) | date | AAAA-MM-DD | 10 | Dfp |
| | mes_doc | Mês do documento (aspecto temporal) | varchar | Númerico | 4 | Dfp |
| | ano_doc | Ano do documento (aspecto temporal) | varchar | Númerico | 4 | Dfp |

Demonstração das Mutações do Patrimônio Líquido (DMPL)

| Arquivo de Origem | Periodicidade de atualização |
|------------------------------|------------------------------|
| meta_dfp_cia_aberta_bdt.zip | Semanal |
| meta_dfp_cia_aberta_DMPL.txt | |

| Colunas Origem | Colunas Destino | Descrição coluna | Tipo de dados | Dominio | Tamanho/Precisão | Tabela de Destino |
|----------------|-----------------------|-----------------------------------------------------------------------------------|---------------|--------------|------------------|-------------------|
| CD_CONTA | codigo_conta | Código da conta | varchar | Númerico | 18 | Dfp |
| CD_CVM | codigo_cvm | Código CVM | char | Númerico | 6 | Dfp |
| CNPJ_CIA | cnpj_companhia | CNPJ da companhia | varchar | Alfanumérico | 20 | |
| COLUMA_DF | coluna_df | Coluna da demonstração financeira | varchar | Alfanumérico | 60 | Dfp |
| DENOM_CIA | denominacao_companhia | Nome empresarial da companhia | varchar | Alfanumérico | 100 | |
| DS_CONTA | descricao_cotna | Descrição da conta | varchar | Alfanumérico | 100 | Dfp |
| DT_FIM_EXERC | data_fim_exercicio | Data fim do exercicio social | date | AAAA-MM-DD | 10 | Dfp |
| DT_INI_EXERC | data_inicio_exercicio | Data inicio do exercicio social | date | AAAA-MM-DD | 10 | Dfp |
| DT_REFER | data_referencia_doc | Data de referência do documento | date | AAAA-MM-DD | 10 | Dfp |
| ESCALA_MOEDA | escala_monetaria | Escala monetária | varchar | Alfanumérico | 100 | Dfp |
| GRUPO_DFP | grupo_dfp | Nome e nível de agregação da demonstração (Mostra se é Individual ou Consolidado) | varchar | Alfanumérico | 206 | grupo_dfp |
| MOEDA | moeda | Moeda | varchar | Alfanumérico | 100 | Dfp |
| ORDEM_EXERC | ordem_exercicio | Ordem do exercicio social | varchar | Alfanumérico | 9 | Dfp |
| ST_CONTA_FIXA | conta_fixa | Indica se é conta fixa ou não | varchar | S/N | 1 | Dfp |
| VERSAO | versao | Versão do documento | smallint | Númerico | 5 | Dfp |
| VL_CONTA | valor_conta | Valor da conta | decimal | Númerico | 29 | Dfp |
| | data_doc | Data do documento (aspecto temporal) | date | AAAA-MM-DD | 10 | Dfp |
| | mes_doc | Mês do documento (aspecto temporal) | varchar | Númerico | 4 | Dfp |
| | ano_doc | Ano do documento (aspecto temporal) | varchar | Númerico | 4 | Dfp |

| | | | | | | |
|---------------|-----------------|--------------------------------------|----------|--------------|----|-----|
| ORDEM_EXERC | ordem_exercicio | Ordem do exercício social | varchar | Alfanumérico | 9 | DfP |
| ST_CONTA_FIXA | conta_fixa | Indica se é conta fixa ou não | varchar | S/N | 1 | DfP |
| VERSAO | versao | Versão do documento | smallint | Número | 5 | DfP |
| VL_CONTA | valor_conta | Valor da conta | decimal | Número | 29 | DfP |
| | data_doc | Data do documento (aspecto temporal) | date | AAAA-MM-DD | 10 | DfP |
| | mes_doc | Mês do documento (aspecto temporal) | varchar | Número | 4 | DfP |
| | ano_doc | Ano do documento (aspecto temporal) | varchar | Número | 4 | DfP |

| Demonstração de Valor Adicionado (DVA) | |
|----------------------------------------|------------------------------|
| Arquivo de Origem | Periodicidade de atualização |
| meta_dfp_cia_aberta_bt.zip | Semanal |
| meta_dfp_cia_aberta_DVA.txt | |

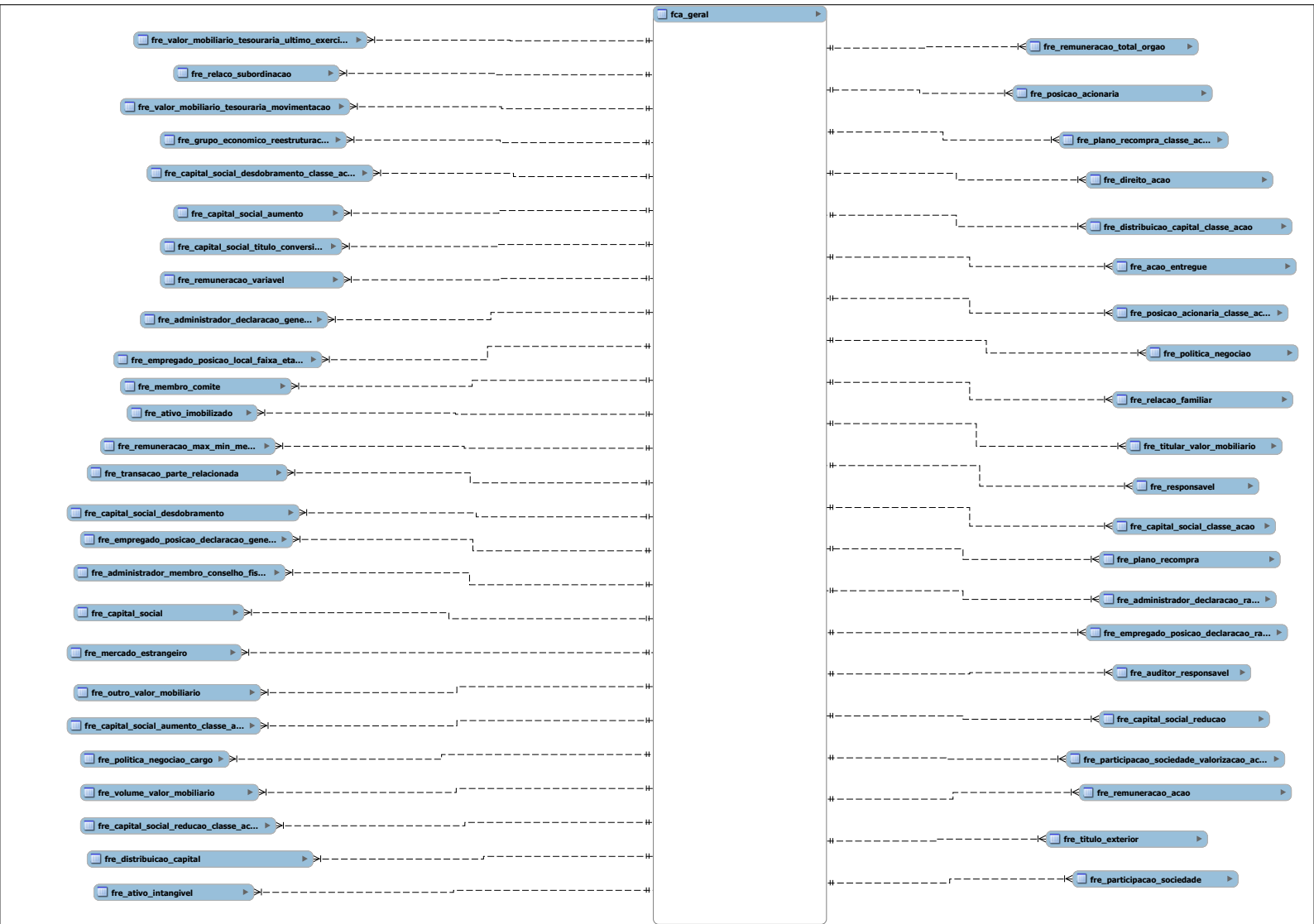
| Colunas Origem | Colunas Destino | Descrição coluna | Tipo de dados | Domínio | Tamanho/Precisão | Tabela de Destino |
|----------------|-----------------------|-----------------------------------------------------------------------------------|---------------|--------------|------------------|-------------------|
| CD_CONTA | codigo_conta | Código da conta | varchar | Número | 18 | Dfp |
| CD_CVM | codigo_cvm | Código CVM | char | Número | 6 | Dfp |
| CNPJ_CIA | cnnpj_companhia | CNPJ da companhia | varchar | Alfanumérico | 20 | |
| DENOM_CIA | denominacao_companhia | Nome empresarial da companhia | varchar | Alfanumérico | 100 | |
| DS_CONTA | descricao_conta | Descrição da conta | varchar | Alfanumérico | 100 | |
| DT_FIM_EXERC | data_fim_exercicio | Data fim do exercício social | date | AAAA-MM-DD | 10 | Dfp |
| DT_INI_EXERC | data_inicio_exercicio | Data início do exercício social | date | AAAA-MM-DD | 10 | Dfp |
| DT_REFER | data_referencia_doc | Data de referência do documento | date | AAAA-MM-DD | 10 | Dfp |
| ESCALA_MOEDA | escala_monetaria | Escala monetária | varchar | Alfanumérico | 100 | |
| GRUPO_DFP | grupo_dfp | Nome e nível de agregação da demonstração (Mostra se é Individual ou Consolidado) | varchar | Alfanumérico | 206 | grupo_dfp |
| MOEDA | moeda | Moeda | varchar | Alfanumérico | 100 | Dfp |
| ORDEM_EXERC | ordem_exercicio | Ordem do exercício social | varchar | Alfanumérico | 9 | Dfp |
| ST_CONTA_FIXA | conta_fixa | Indica se é conta fixa ou não | varchar | S/N | 1 | Dfp |
| VERSAO | versao | Versão do documento | smallint | Número | 5 | Dfp |
| VL_CONTA | valor_conta | Valor da conta | decimal | Número | 29 | Dfp |
| | data_doc | Data do documento (aspecto temporal) | date | AAAA-MM-DD | 10 | Dfp |
| | mes_doc | Mês do documento (aspecto temporal) | varchar | Número | 4 | Dfp |
| | ano_doc | Ano do documento (aspecto temporal) | varchar | Número | 4 | Dfp |

[illegible]

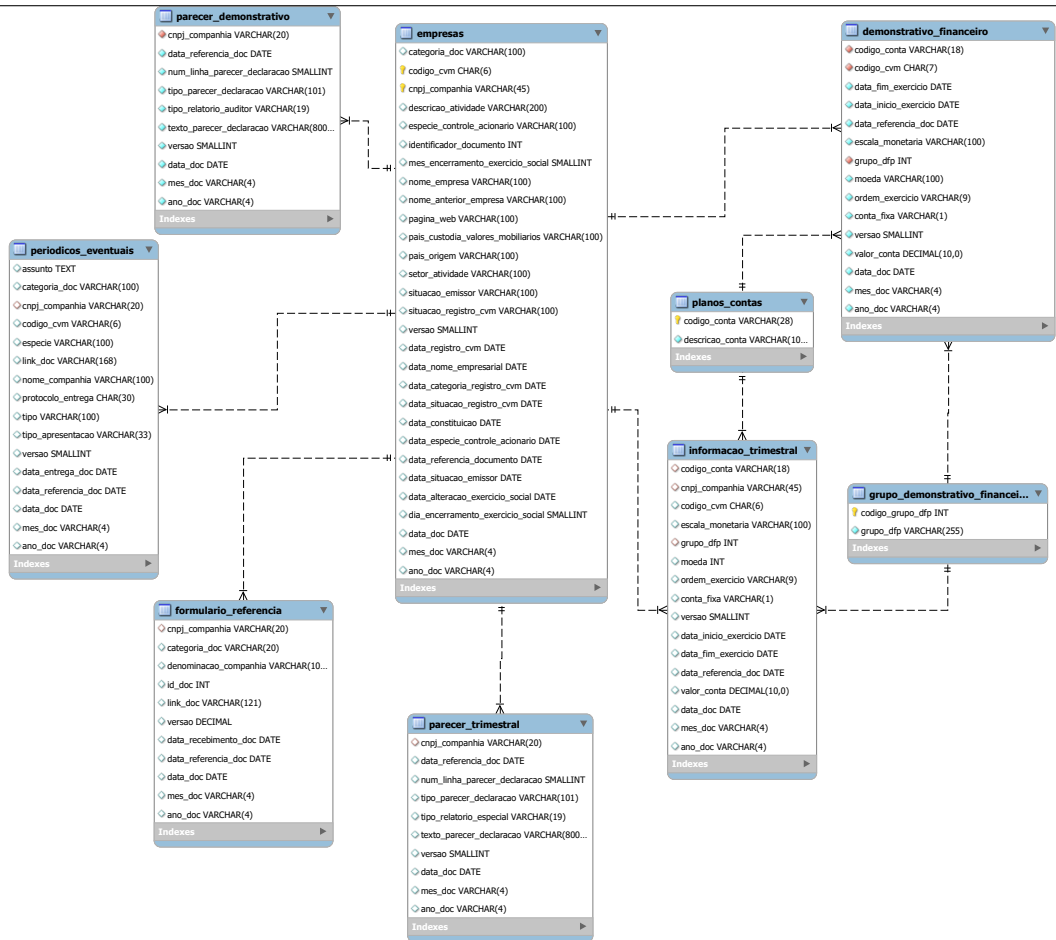
| Colunas Origem | Colunas Destino | Descrição coluna | Tipo de dados | Domínio | Tamanho/Precisão | Tabela de Destino |
|-----------------------|------------------------------|------------------------------------------------|---------------|--------------|------------------|-------------------|
| CNPJ_CIA | cnpj_companhia | CNPJ da companhia | varchar | Alfanumérico | 20 | Dfp_Parecer |
| DENOM_CIA | denominacao_companhia | Nome empresarial da companhia | varchar | Alfanumérico | 100 | |
| DT_REFER | data_referencia_doc | Data de referência do documento | date | AAAA-MM-DD | 10 | Dfp_Parecer |
| NUM_ITEM_PARECER_DECL | num_linha_parecer_declaracao | Número da linha do texto do Parecer/Declaração | smallint | Número | 5 | Dfp_Parecer |

| | | | | | | |
|-------------------|--------------------------|-------------------------------------------------------------|----------|--------------|------|-------------|
| TP_PARECER_DECL. | tipo_parecer_declaracao | Tipo do Parecer/Declaração | varchar | Alfanumérico | 101 | Dfp_Parecer |
| TP_RELAT_AUD | tipo_relatorio_auditor | Tipo Relatório do Auditor Independente | varchar | Alfanumérico | 19 | Dfp_Parecer |
| TXT_PARECER_DECL. | texto_parecer_declaracao | Texto do Parecer/Declaração | varchar | Alfanumérico | 8000 | Dfp_Parecer |
| VERSAO | versao | Versão do documento | smallint | Númérico | 5 | Dfp_Parecer |
| | data_doc | Data do documento (aspecto temporal) | date | AAAA-MM-DD | 10 | Dfp_Parecer |
| | mes_doc | Mês do documento (aspecto temporal) | varchar | Númérico | 4 | Dfp_Parecer |
| | ano_doc | Ano do documento (aspecto temporal) | varchar | Númérico | 4 | Dfp_Parecer |
| | | | | | | |
| | | | | | | |
| grupo_dfp | | | | | | |
| codigo_grupo_dfp | int | índice da tabela para agrupar os grupos de DPF | | | | |
| | varchar(255) | texto que mostra se individual ou consolidado de cada grupo | | | | |

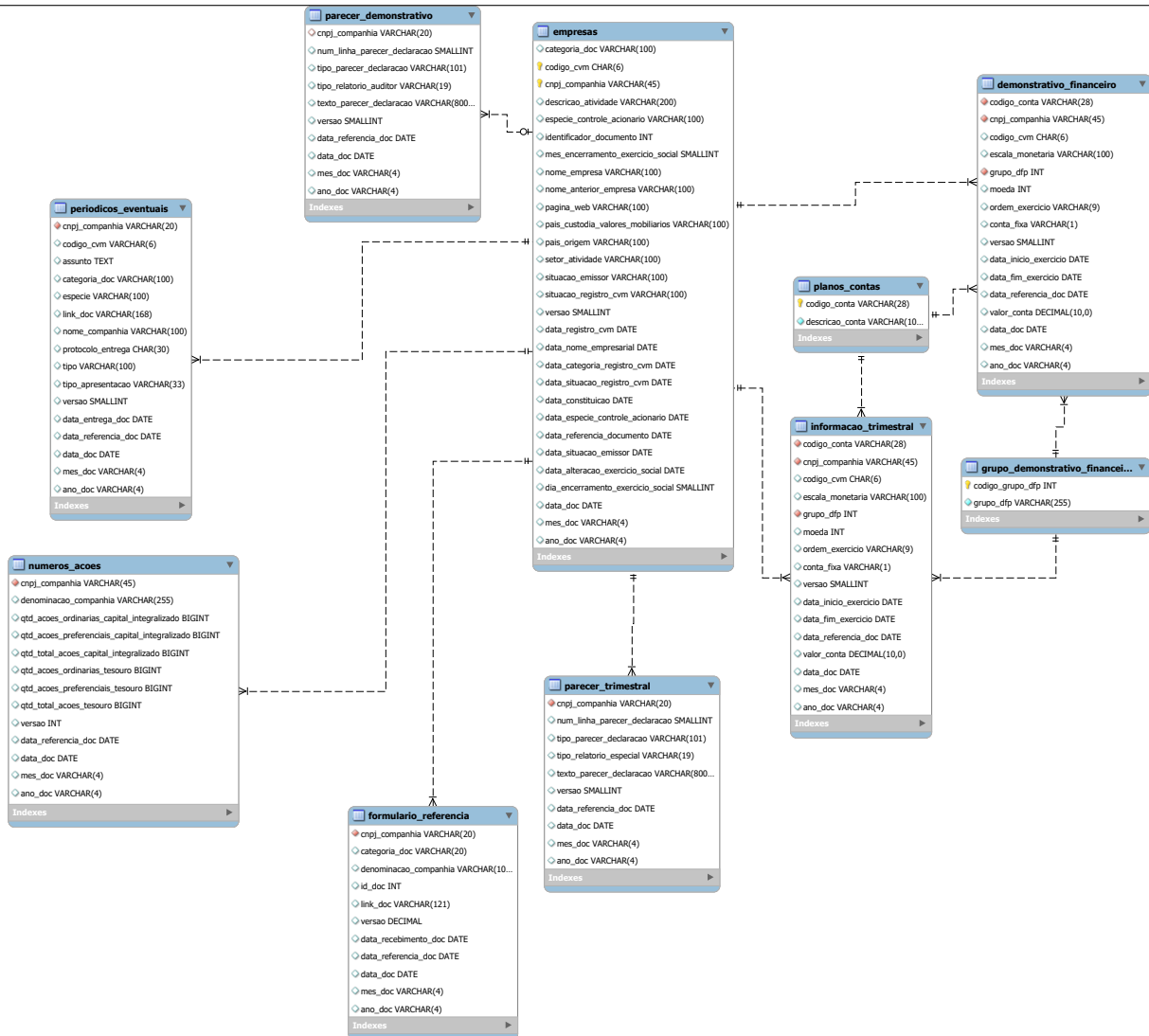
APÊNDICE H - PRIMEIRA MODELAGEM



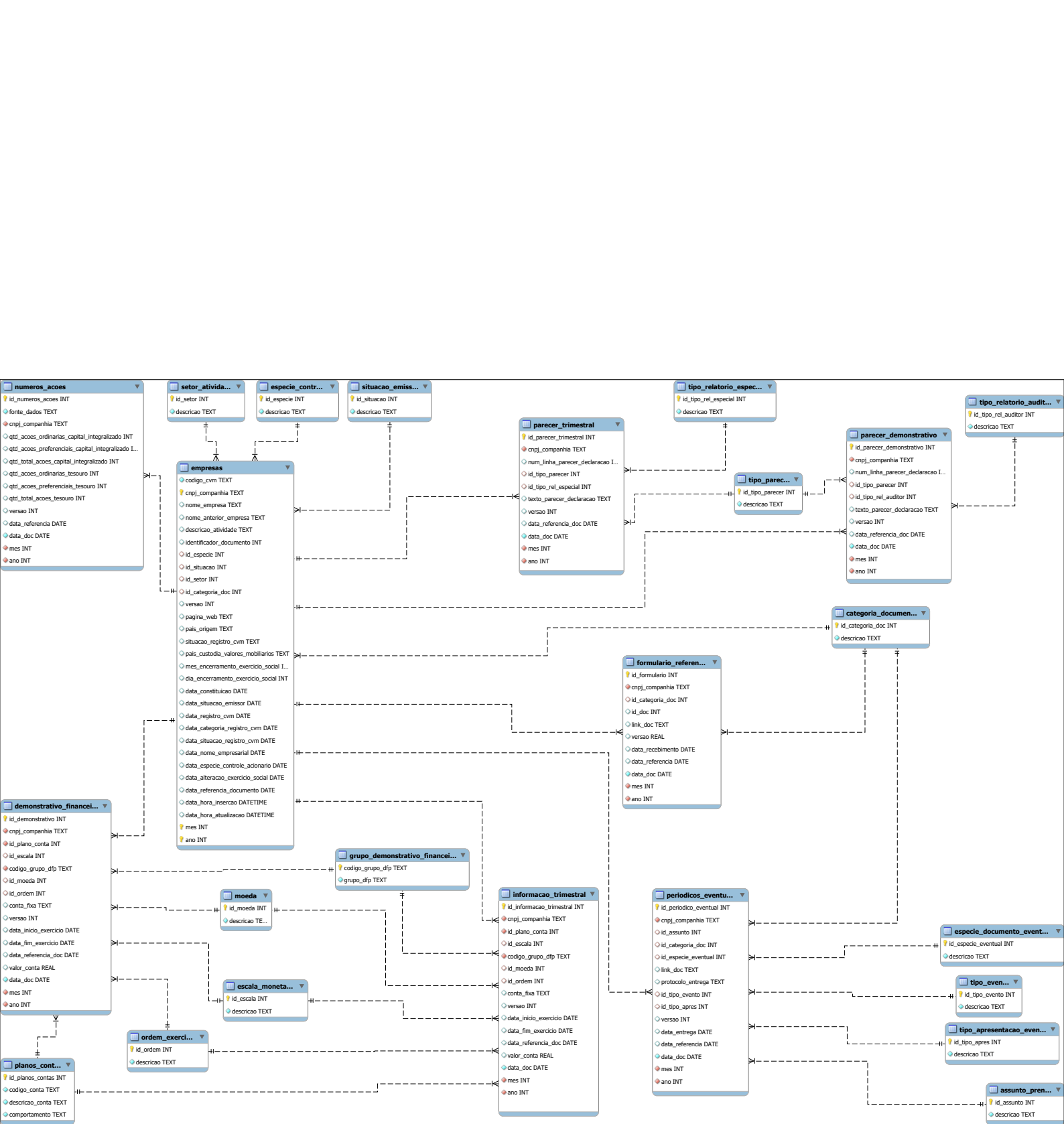
APÊNDICE I - MODELAGEM APOS AJUSTES INICIAIS



APÊNDICE J - PENÚLTIMA MODELAGEM



APÊNDICE K - MODELAGEM PRESENTE NO CÓDIGO



ÍNDICE

INTRODUÇÃO, 15