

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**Data Warehouse para transparência
orçamentária municipal: uma proposta para
Juiz de Fora**

Antônio Marcos Souza Pereira

Data Warehouse para transparência orçamentária municipal: uma proposta para Juiz de Fora

ANTÔNIO MARCOS SOUZA PEREIRA

Universidade Federal de Juiz de Fora

Instituto de Ciências Exatas

Departamento de Ciência da Computação

Bacharelado em Ciência da Computação

Orientador: Victor Ströele de Andrade Menezes

JUIZ DE FORA

2026

DATA WAREHOUSE PARA TRANSPARÊNCIA ORÇAMENTÁRIA MUNICIPAL: UMA PROPOSTA PARA JUIZ DE FORA

Antônio Marcos Souza Pereira

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS
EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTE-
GRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE
BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Victor Ströele de Andrade Menezes
Doutor em Engenharia de Sistemas e Computação pela UFRJ

José Maria Nazar David
Doutor em Engenharia de Sistemas e Computação pela UFRJ

Regina Maria Maciel Braga Villela
Doutora em Engenharia de Sistemas e Computação pela UFRJ

JUIZ DE FORA

2026

Resumo

A divulgação de dados orçamentários municipais avançou com a ampliação dos Portais da Transparência, mas a forma como essas informações são disponibilizadas ainda impõe barreiras à compreensão da população. Planilhas extensas, formatos heterogêneos e linguagem técnica dificultam o acompanhamento efetivo da execução orçamentária e limitam o exercício do controle social. Este trabalho propõe e implementa um *Data Warehouse* para análise da execução orçamentária de Juiz de Fora, integrando informações dispersas do Portal da Transparência municipal, do ementário da Secretaria do Tesouro Nacional e da Lei Orçamentária Anual em um modelo dimensional. A pesquisa adota *Design Science Research* como abordagem metodológica e a modelagem dimensional de Kimball como referência técnica. O artefato resultante integra dez dimensões e três tabelas de fatos de despesas e receitas, alimentadas por um *pipeline* de ETL idempotente que documenta explicitamente as inconsistências encontradas nas fontes. O consumo ocorre por meio de sete painéis interativos publicados em instância pública, avaliados comparativamente ao portal original em tarefas representativas de fiscalização. Os resultados mostram uma redução consistente do esforço de consolidação manual exigido do usuário e revelam padrões recorrentes na execução orçamentária municipal, entre eles a dependência de transferências intergovernamentais e a distância entre o empenho e o pagamento efetivo.

Palavras-chave: transparência pública; orçamento municipal; *data warehouse*; modelagem dimensional; *design science research*; controle social.

Abstract

The disclosure of municipal budget data has advanced with the expansion of Transparency Portals, but the way this information is made available still imposes barriers to public understanding. Extensive spreadsheets, heterogeneous formats, and technical language hinder effective monitoring of budget execution and limit the exercise of social control. This work proposes and implements a Data Warehouse to analyze budget execution in Juiz de Fora, integrating dispersed information from the municipal Transparency Portal, the National Treasury Secretariat's revenue classification catalog, and the Annual Budget Law into a dimensional model. The research adopts Design Science Research as its methodological approach and Kimball's dimensional modeling as its technical reference. The resulting artifact integrates ten dimensions and three fact tables on expenditures and revenues, fed by an idempotent ETL pipeline that explicitly documents the inconsistencies found in the source data. Consumption is provided through seven interactive dashboards published on a public instance and evaluated against the original portal for representative fiscalization tasks. The results show a consistent reduction in the manual consolidation effort required of the user and reveal recurring patterns in municipal budget execution, including dependence on intergovernmental transfers and a gap between commitment and effective payment.

Keywords: public transparency; municipal budget; data warehouse; dimensional modeling; design science research; social control.

Agradecimentos

À minha mãe, por todos os sacrifícios que me trouxeram até aqui. Aos meus amigos e à minha namorada, pela companhia e pelo apoio durante essa caminhada. Ao meu orientador, pelo apoio nesta monografia. A todos os funcionários do Departamento de Ciência da Computação da UFJF, que, de alguma forma, contribuíram com a minha formação até aqui. E ao povo brasileiro, que financiou cada dia desta graduação. Universidade pública, gratuita e de qualidade é conquista popular, e é dever de todos defendê-la.

*"A leitura do mundo precede a leitura da
palavra".*

Paulo Freire

Conteúdo

Lista de Figuras	7
Lista de Tabelas	8
Lista de Abreviações	9
1 Introdução	10
1.1 Problema de Pesquisa	12
1.2 Objetivos	12
1.3 Delimitação do Escopo	13
1.4 Estrutura do Trabalho	13
2 Fundamentação Teórica	15
2.1 Orçamento Público e Transparência Governamental	15
2.2 Controle Social e Transparência	16
2.3 Data Warehouse e Modelagem Dimensional	17
3 Trabalhos Relacionados	19
3.1 Avaliação de Portais de Transparência Municipais	19
3.2 Data Warehouse e Análise de Dados Governamentais	20
3.3 <i>Civic Tech</i> e Vigilância Civil	23
3.4 Síntese e Posicionamento deste Trabalho	23
4 Metodologia	25
4.1 Abordagem Metodológica: Design Science Research	25
4.2 Revisão Bibliográfica	26
4.3 Pesquisa Documental e Seleção de Fontes	27
4.4 Desenvolvimento Iterativo do Artefato	28
4.5 Documentação de Inconsistências	29
4.6 Verificação por Invariantes de Domínio	29
4.7 Avaliação Funcional Comparativa	30
4.8 Compromisso com Reprodutibilidade	32
4.9 Cenário de Uso: Juiz de Fora	32
5 Modelagem e Implementação do Data Warehouse	34
5.1 Visão Geral da Arquitetura	34
5.2 Tecnologias Utilizadas	36
5.3 Extração: Portal da Transparência, STN e LOA PJF	37
5.3.1 Detecção Dinâmica de Cabeçalho	37
5.3.2 Quebra de <i>Layout</i> em 2025-05	38
5.3.3 Classificação Funcional da LOA (exercício 2026)	38
5.3.4 Idempotência por Estratégia de Escrita	39
5.4 Camada de Staging	39
5.5 Modelo Dimensional	39
5.5.1 Tabelas de Fatos	40

5.5.2	Dimensões	40
5.5.3	Chaves Substitutas	41
5.5.4	Esquema Estrela	42
5.6	Processos de ETL e Idempotência	43
5.7	Qualidade de Dados e Testes	43
5.8	Limitações da Implementação	44
5.8.1	Granularidade da Despesa	44
5.8.2	Ausência de Orçamento Autorizado	44
5.8.3	Classificação MCASP Funcional Incompleta	45
6	Resultados: Dashboards e Indicadores	46
6.1	Visão Geral da Execução Orçamentária	46
6.2	Execução por Função de Governo	47
6.3	Concentração de Fornecedores	49
6.4	Receitas Municipais	50
6.5	Análise dos Dados de Juiz de Fora	51
7	Avaliação e Discussão	54
7.1	Avaliação Funcional	54
7.2	Limitações do Trabalho	54
7.3	Implicações para Transparência e Controle Social	56
7.4	Reprodutibilidade, Replicabilidade e Operação	56
8	Conclusão	58
8.1	Síntese das Contribuições	58
8.2	Considerações Finais	59
8.3	Trabalhos Futuros	59
	Bibliografia	61

Lista de Figuras

5.1	Visão conceitual da arquitetura implementada. Os blocos tracejados indicam o ponto de extensão para uma nova fonte; os demais blocos mostram a instanciãção usada no cenário de Juiz de Fora.	35
5.2	Esquemas estrela do <i>warehouse</i> , compartilhando <code>dim_tempo</code>	42
5.3	Modelo dimensional do <i>warehouse</i> : três tabelas de fatos articuladas por dimensões conformadas. <code>dim_tempo</code> é compartilhada pelas três fatos e atua em dois papéis (<i>role-playing</i>) na estrela de despesa. <code>dim_natureza_receita</code> é conformada entre <code>fct_receita</code> e <code>fct_receita_acumulada</code>	42
6.1	Visão Geral Fiscal: resumo do recorte e comparação do último mês (seções 1 e 2).	47
6.2	Visão Geral Fiscal: evolução mensal de arrecadação <i>vs.</i> pagamentos (seção 3).	48
6.3	Despesas: resumo dos estágios da execução e da evolução mensal empenhada, liquidada e paga.	49
6.4	Execução Orçamentária por função e subfunção: gráfico das vinte maiores combinações e tabela funcional.	50
6.5	Fornecedores: concentração de pagamentos (top 20) e ranking completo com participação percentual.	51
6.6	Receitas: resumo do recorte e composição por origem.	52

Lista de Tabelas

3.1	Posicionamento comparativo em relação à literatura revisada.	24
5.1	<i>Stack</i> técnica utilizada no DW	36
5.2	Fontes de dados consumidas pelo <i>pipeline</i> de extração	38
5.3	Tabelas de fatos do <i>warehouse</i>	40
5.4	Dimensões do <i>warehouse</i>	41
5.5	Testes singulares de qualidade de dados	44
7.1	Comparação portal PJF <i>versus dashboard</i> em tarefas de fiscalização	55

Lista de Abreviações

API	<i>Application Programming Interface</i>
BI	<i>Business Intelligence</i>
CGU	Controladoria-Geral da União
DCC	Departamento de Ciência da Computação
DW	<i>Data Warehouse</i>
ETL	<i>Extract, Transform, Load</i>
LAI	Lei de Acesso à Informação
LDO	Lei de Diretrizes Orçamentárias
LOA	Lei Orçamentária Anual
MCASP	Manual de Contabilidade Aplicada ao Setor Público
MTO	Manual Técnico de Orçamento
OCR	<i>Optical Character Recognition</i>
PJF	Prefeitura de Juiz de Fora
PPA	Plano Plurianual
SCD	<i>Slowly Changing Dimension</i>
SQL	<i>Structured Query Language</i>
STN	Secretaria do Tesouro Nacional
UFJF	Universidade Federal de Juiz de Fora

1 Introdução

A Lei de Acesso à Informação (LAI), promulgada em 2011, assegurou ao cidadão o direito de obter informações produzidas pelo Estado brasileiro (BRASIL, 2011). Os portais municipais são um dos meios usados para divulgar dados sobre orçamento, finanças e administração. O acesso aos arquivos, contudo, não garante sua interpretação. Classificações contábeis, tabelas separadas e poucos recursos de análise ainda dificultam o acompanhamento dos gastos por usuários não especializados (MATIAS-PEREIRA, 2010; MELO; FUCHIGAMI, 2019).

Um Data Warehouse (DW) integra fontes heterogêneas e organiza os dados para consulta analítica (KIMBALL; ROSS, 2013). Neste trabalho, as perguntas de análise foram definidas a partir das dificuldades de cruzamento relatadas por organizações de fiscalização (MARCO; TERCI, 2022) e dos campos disponíveis nas fontes de Juiz de Fora. Entre elas estão: como evoluíram os pagamentos classificados na função Saúde; qual parcela da receita prevista foi arrecadada; e quanto os cinco maiores fornecedores representam no total pago. Essas perguntas orientaram a escolha dos fatos, das dimensões e das visualizações. Elas não resultam de entrevistas com usuários.

Para organizar o problema de pesquisa, a distância entre disponibilidade e uso dos dados é examinada sob três aspectos: social, técnico e pedagógico. Essa divisão é uma síntese adotada neste trabalho. O aspecto social decorre das barreiras enfrentadas por cidadãos e organizações de fiscalização (MARCO; TERCI, 2022); o técnico, das limitações observadas nos portais (MELO; FUCHIGAMI, 2019); e o pedagógico, da discussão sobre leitura crítica e conscientização (FREIRE, 1967; FREIRE, 1996).

O aspecto social envolve o conhecimento necessário para interpretar o orçamento e relacioná-lo a serviços públicos concretos. Categorias como elemento de despesa, funcional programática e fonte de recurso não correspondem diretamente à linguagem usada no cotidiano. Mesmo os Observatórios Sociais estudados por Marco e Terci (2022) relatam dificuldades para cruzar informações municipais. Nesse contexto, o acesso formal pode continuar dependente de pessoas ou organizações com conhecimento técnico.

O aspecto técnico diz respeito à forma de publicação. No portal analisado, receitas e despesas são distribuídas em planilhas distintas, os cabeçalhos não ocupam sempre a mesma posição e houve mudança de *layout* durante o período estudado. Uma análise histórica exige baixar os arquivos mensais, localizar as tabelas, uniformizar colunas, interpretar códigos e combinar os períodos. O DW executa essas etapas no *pipeline*, registra as regras de transformação e mantém o resultado em um modelo que pode ser consultado novamente.

A lacuna pedagógica é a menos debatida e fundamenta-se na obra de Paulo Freire, em especial na noção de educação libertadora e no conceito de conscientização (FREIRE, 1967). Freire argumenta que a transformação social depende da capacidade de "ler o mundo" criticamente, superando a passividade diante das estruturas de dominação. Aplicado ao orçamento público, isso significa que a democratização do poder vai além da publicação passiva de dados: exige instrumentos que capacitem o cidadão a formular perguntas, testar hipóteses e construir suas próprias leituras sobre a gestão pública (FREIRE, 1996). O conceito freireano de "inédito viável" descreve com precisão o lugar do Data Warehouse civil. Trata-se de uma forma de participação ainda incipiente, em que o cidadão questiona não só quanto foi gasto, mas também por que aquela área recebeu mais recursos, se o investimento responde às necessidades locais e se há sinais de favorecimento a fornecedores específicos.

A execução orçamentária municipal foi escolhida porque decisões locais afetam serviços próximos ao cotidiano dos moradores e dispõem de mecanismos de participação direta (NOVAES; SANTOS, 2014). O trabalho documenta uma implementação em Juiz de Fora; sua adaptação a outros municípios depende dos formatos e campos publicados por cada portal. O recorte combina a discussão sobre transparência pública com a modelagem de dados. Nessa combinação, a arquitetura resulta de prioridades organizacionais que orientam a seleção das fontes, as classificações expostas e as consultas oferecidas ao público.

1.1 Problema de Pesquisa

A pergunta que orienta este trabalho é: *Como organizar, em um modelo dimensional, os dados publicados pelo Portal da Transparência da Prefeitura de Juiz de Fora, de forma a reduzir as barreiras técnicas que hoje afastam cidadãos não especialistas da execução orçamentária municipal?*

No plano técnico, a pergunta exige definir o grão dos fatos, mapear as classificações oficiais, tratar inconsistências e permitir a repetição das cargas. No plano de uso, exige escolher rótulos, indicadores e formas de navegação compreensíveis sem conhecimento prévio do esquema das planilhas. Essas decisões são documentadas e avaliadas nos capítulos seguintes.

1.2 Objetivos

Objetivo Geral

Desenvolver e avaliar um DW para a execução orçamentária de Juiz de Fora. A dimensão tecnológica do artefato integra e modela os dados publicados; a dimensão didática oferece rótulos, glossário e visualizações que apoiam sua interpretação por cidadãos não especializados.

Objetivos Específicos

- Coletar e documentar dados de receitas e despesas do Portal da Transparência da Prefeitura de Juiz de Fora, identificando inconsistências nas fontes disponíveis.
- Modelar dimensionalmente a execução orçamentária segundo a metodologia Kimball, mapeando as classificações oficiais (MCASP/MTO) em fatos e dimensões.
- Implementar processos de ETL para extração, limpeza, transformação e carga de dados, tratando heterogeneidades de formato e de nomenclatura.
- Construir dashboards interativos que traduzam classificações orçamentárias oficiais em visualizações acessíveis, com indicadores derivados das tabelas de fatos do

warehouse.

- Avaliar, em tarefas representativas de fiscalização, o painel desenvolvido em comparação com o Portal da Transparência da Prefeitura de Juiz de Fora, identificando ganhos e limitações de usabilidade analítica.

1.3 Delimitação do Escopo

Este trabalho concentra-se na execução orçamentária de Juiz de Fora. O recorte geográfico limita a aplicação e a avaliação às fontes do município. A arquitetura pode orientar adaptações para outros portais, mas essa possibilidade não foi verificada empiricamente.

O recorte temático cobre receitas e despesas no nível publicado pelo portal, com dados complementares do Ementário de Classificação por Natureza da Receita da STN e da classificação funcional da LOA. Licitações, contratos, folha de pessoal e patrimônio não são integrados.

Quanto à granularidade, a fonte contém relatórios mensais consolidados. O modelo mantém o grão das linhas publicadas. Análises de transações, itens ou eventos abaixo do empenho exigem outras bases.

A avaliação compara tarefas executadas no painel e no portal original. Não foram realizados testes de usabilidade com participantes. Esse tipo de estudo exigiria protocolo próprio, submissão ao comitê de ética e recrutamento, atividades que não foram incluídas no cronograma do TCC.

1.4 Estrutura do Trabalho

Os sete capítulos seguintes apresentam a fundamentação teórica, os trabalhos relacionados e a metodologia. Em seguida, o Capítulo 5 descreve a arquitetura, o modelo dimensional e o *pipeline*; o Capítulo 6 apresenta o painel e os indicadores; e o Capítulo 7 compara o artefato com o portal municipal. O Capítulo 8 reúne os resultados, as limitações e os trabalhos futuros.

O estudo adota *Design Science Research* (DSR) e executa uma iteração completa,

da definição do problema à avaliação do artefato (HEVNER et al., 2004; DRESCH; LACERDA; JÚNIOR, 2015). A abordagem organiza a relação entre as decisões de projeto, o DW construído e os critérios usados para avaliá-lo. O Capítulo 4 detalha as cinco etapas percorridas nessa iteração.

2 Fundamentação Teórica

Este capítulo reúne os três temas que fundamentam o trabalho. O orçamento público é o objeto de estudo, o controle social define a finalidade da análise e o Data Warehouse fornece o modelo técnico adotado. A estrutura dos dados decorre das prioridades e metas da organização responsável pelo sistema. Essas diretrizes orientam decisões arquiteturais, como a seleção das fontes e a definição das dimensões de análise. Com isso, influenciam as perguntas que podem ser respondidas e o uso dos dados pelo cidadão no controle social.

2.1 Orçamento Público e Transparência Governamental

O orçamento público brasileiro se organiza em três instrumentos integrados: o Plano Plurianual (PPA), a Lei de Diretrizes Orçamentárias (LDO) e a Lei Orçamentária Anual (LOA) (GIACOMONI, 2021). O PPA define metas e diretrizes para um período de 4 anos. A LDO estabelece prioridades para o exercício seguinte. A LOA estima receitas e fixa despesas com base anual. Esse encadeamento confere ao orçamento brasileiro uma característica importante para a análise: as decisões financeiras de um ano específico não nascem isoladas, mas decorrem da execução de planos plurianuais.

As despesas são classificadas em quatro dimensões: institucional (qual órgão executa), funcional (a que função de governo se destina), programática (qual programa atende) e por natureza (tipo de gasto). Cada despesa percorre três estágios distintos antes de virar pagamento efetivo: o empenho reserva o crédito orçamentário, a liquidação verifica o direito do credor e o pagamento efetiva a transferência. A separação entre essas etapas explica fenômenos comuns no debate público, como os "restos a pagar" e as diferenças entre o valor empenhado e o valor pago. Sem entender essa cadeia, é fácil tirar conclusões equivocadas a partir dos dados.

Documentos normativos, como o Manual de Contabilidade Aplicada ao Setor

Público (MCASP) e o Manual Técnico de Orçamento (MTO), padronizam essas classificações em âmbito nacional (BRASIL, 2024a; BRASIL, 2024b). A padronização é uma virtude para o controle interno e para a contabilidade pública, mas se torna um obstáculo quando o destinatário é o cidadão. Pesquisas indicam que termos como "elemento de despesa", "função programática" e "fonte de recurso" atuam como barreiras semânticas e dificultam a interpretação por não especialistas (MELO; FUCHIGAMI, 2019; MARCO; TERCI, 2022). O problema não é a falta de padrão, mas sim um padrão pensado para o emissor, e não para o receptor.

2.2 Controle Social e Transparência

A LAI consolidou a transparência governamental como obrigação legal, mas estudos consistentes mostram que publicar não é o mesmo que prestar contas. Pinho e Sacramento (2009) discute essa distância entre transparência formal e *accountability* efetiva no contexto brasileiro: a primeira tem caráter declarativo, a segunda exige resposta, justificativa e eventual sanção. Há um abismo entre disponibilizar um arquivo CSV no portal e responder à pergunta que o cidadão de fato faria. Em registro próximo, Baldo (2018) sustenta que a democratização do orçamento exige a articulação simultânea de três princípios, a saber, legalidade, legitimidade e economicidade. Privilegiar qualquer um deles, isoladamente, compromete o resultado: legalismo burocrático sem participação real, economicismo que entrega o espaço público a interesses privados ou legitimidade sem ancoragem que gera expectativas inviáveis.

A participação cidadã qualificada esbarra na linguagem técnico-burocrática. Marco e Terceiro (2022) mostram, em estudo com Observatórios Sociais, que mesmo organizações civis dedicadas à fiscalização enfrentam dificuldade para cruzar dados de portais municipais. Se grupos organizados encontram esse obstáculo, o cidadão isolado dificilmente terá condições melhores.

Paulo Freire oferece um vocabulário próprio para discutir esse problema. Sua noção de conscientização aponta para a capacidade de "ler o mundo" criticamente, indo além da decodificação passiva de signos (FREIRE, 1967; FREIRE, 1996). Streck, Pitano e Moretti (2017) investigam empiricamente a presença desse legado em experiências de

orçamento participativo no Rio Grande do Sul e mostram que as formulações inspiradas em Freire convivem com fatores que limitam seu potencial formativo. A intenção pedagógica das políticas participativas, na prática, costuma se descolar do que de fato chega ao cidadão. Aplicada ao orçamento público, a ideia ganha contornos técnicos. A conscientização orçamentária implica acessar os dados, compreender o significado dos termos, identificar padrões ao longo do tempo e formular perguntas próprias. A organização da informação não é neutra. Ela define o universo de leitores possíveis e, com isso, o universo de fiscalizadores possíveis.

2.3 Data Warehouse e Modelagem Dimensional

O Data Warehouse (DW) é um repositório voltado à análise, e Kimball e Ross (2013) resumem suas propriedades em uma definição clássica: orientado a assunto, integrado, não volátil e variável no tempo. Cada propriedade tem efeito prático no uso analítico.

Orientado a assunto significa que os dados são organizados em torno de temas relevantes para análise, como despesas e receitas, e não em torno de processos transacionais como "cadastro de empenho". A diferença entre as duas organizações se torna evidente quando o analista pergunta "quanto a saúde gastou em 2024". Em um sistema transacional, essa pergunta exige a junção de dezenas de tabelas. Em um DW, ela cabe em uma consulta de poucas linhas.

Integrado quer dizer que dados de fontes heterogêneas chegam ao DW sob convenções uniformes de nome, tipo, unidade e formato. No caso público, isso resolve um problema concreto: arquivos de anos diferentes do mesmo portal costumam usar nomenclaturas distintas para o mesmo conceito. A integração elimina o esforço de reconciliação manual que, hoje, qualquer usuário sério dos portais é obrigado a realizar.

Não volátil indica que os dados carregados não sofrem atualizações destrutivas, apenas se acrescentam novos dados. Essa propriedade preserva o histórico e confere ao DW a característica de arquivo público auditável, em que nenhuma versão antiga é substituída silenciosamente.

Variável no tempo significa que o histórico é tratado como uma dimensão analítica. Cada fato carrega sua marcação temporal. Isso permite análises que comparam anos, iden-

tificam sazonalidades e detectam mudanças de padrão. Para a execução orçamentária, é justamente aí que residem as perguntas mais relevantes.

A modelagem dimensional, consolidada por Kimball e Ross, organiza o DW em torno do esquema em estrela. Tabelas de fatos armazenam métricas quantitativas (no caso orçamentário, valor empenhado, valor liquidado e valor pago). Tabelas de dimensão fornecem contexto descritivo: tempo, órgão, função, programa, natureza da despesa, favorecido, fonte de recurso. Essa arquitetura prioriza a usabilidade analítica e dispensa o usuário de dominar as junções complexas típicas de modelos relacionais normalizados.

Neste trabalho, as duas referências são usadas em planos distintos. A modelagem dimensional de Kimball orienta a organização técnica dos dados, enquanto a obra de Freire fornece a base para discutir sua apropriação pelo cidadão. Os *dashboards* construídos sobre o DW podem reunir séries históricas e comparações entre áreas que antes exigiam o tratamento manual de planilhas. Sua contribuição para a transparência depende de o cidadão compreender os conceitos apresentados e conseguir elaborar perguntas sobre a execução orçamentária. Assim, Kimball fundamenta a estrutura analítica, e Freire, a discussão sobre seu uso no controle social.

3 Trabalhos Relacionados

Este capítulo examina trabalhos sobre avaliação de portais públicos, uso de *Data Warehouse* em dados governamentais e iniciativas de *civic tech*. A comparação considera a esfera de governo atendida, a integração das fontes, o tratamento de inconsistências e os recursos oferecidos ao usuário. Esses critérios permitem situar o artefato desenvolvido nesta pesquisa em relação às soluções e aos estudos existentes.

3.1 Avaliação de Portais de Transparência Municipais

O esforço para medir a qualidade dos portais de transparência brasileiros não é novo. Melo e Fuchigami (2019) propõem um índice bidimensional de transparência da informação público-eletrônica, dividido em transparência ativa (o que o portal disponibiliza espontaneamente) e transparência passiva (a resposta a pedidos formais de informação). Os autores aplicam o índice a municípios brasileiros e mostram que a maior parte deles apresenta níveis baixos a moderados, com forte heterogeneidade entre as regiões. O resultado importa porque revela uma assimetria estrutural na qual o cidadão de um município pequeno do interior tem, em média, menos acesso à informação do que o cidadão de uma capital. A LAI é a mesma, mas a implementação não.

Marco e Terci (2022) adotam uma abordagem complementar e investigam a transparência municipal sob a perspectiva dos Observatórios Sociais. Em vez de medir o portal do lado da oferta, os autores ouvem o lado da demanda, entrevistando organizações civis que efetivamente usam essas plataformas para fiscalização. O achado é consistente com o argumento freireano de que mesmo grupos organizados, com conhecimento técnico médio, têm dificuldades para cruzar dados de despesas, licitações e contratos. A barreira de uso permanece, ainda que o usuário seja qualificado. Se o cidadão organizado tropeça, o cidadão comum sequer começa a caminhar.

Pinho e Sacramento (2009) oferecem o pano de fundo teórico para essas avaliações empíricas. Ao discutirem se “*accountability*” tem tradução adequada para o português, os autores expõem a distância entre publicar informação e prestar contas de fato. A transparência é condição necessária, mas não suficiente, para a *accountability* democrática. Esse referencial dá nome ao problema que este trabalho aborda e justifica por que a simples publicação de dados em portais não fecha o ciclo da participação cidadã.

Teles e Carvalho (2024) sintetizam trinta e cinco anos de produção acadêmica brasileira sobre transparência pública municipal em uma revisão de escopo que abrange 336 trabalhos, entre artigos, dissertações e teses. A revisão mostra concentração das pesquisas em transparência digital ativa e em municípios das regiões Sul e Sudeste, com escassez de estudos técnicos sobre *como* viabilizar a interpretação dos dados publicados. Esse achado é particularmente relevante para este trabalho, pois relata que a literatura brasileira documenta amplamente as *falhas* dos portais, mas oferece poucos exemplos concretos de *soluções técnicas* para o problema.

Em registro complementar, Souza, Miranda e Farias (2025) reúnem evidência recente sobre transparência e controle social no contexto do orçamento participativo brasileiro. Os autores identificam três obstáculos persistentes à participação qualificada: baixa escolaridade política, descontinuidade administrativa entre gestões e apropriação estratégica da participação pelo próprio poder público, fenômeno em que mecanismos formalmente participativos são utilizados para legitimar decisões já tomadas. A leitura conjunta com os trabalhos anteriores mostra que o problema da transparência efetiva não vem sendo resolvido com o passar do tempo, apesar dos avanços legais e do amadurecimento dos portais.

3.2 Data Warehouse e Análise de Dados Governamentais

A aplicação de *Data Warehouse* a dados públicos já é abordada na literatura brasileira, mas de forma pontual e dispersa. A referência canônica da área é *The Data Warehouse Toolkit*, livro originalmente publicado por Ralph Kimball nos Estados Unidos em 1996

(KIMBALL; ROSS, 2013). A obra consolidou a modelagem dimensional como metodologia padrão para projetos de *Business Intelligence* em escala internacional. Os exemplos discutidos pelos autores partem, em sua maioria, de aplicações comerciais, como o varejo, o transporte, as telecomunicações e a gestão financeira corporativa. Esse recorte original se reflete na produção nacional, na qual Almeida (2004), em um dos primeiros trabalhos brasileiros sobre a viabilidade técnica de *Data Warehouse*, discute uma plataforma de baixo custo, sem tratamento específico, voltada ao setor público. Trabalhos posteriores mantiveram o foco em aplicações empresariais.

Cardoso (2019) desenvolve um método para processar Dados Governamentais Abertos dos poderes Legislativo e Executivo federais com técnicas de ciência de dados e ferramentas analíticas. O estudo usa fontes e métodos diferentes dos adotados nesta pesquisa: trabalha com dados abertos federais e não aplica a modelagem dimensional de Kimball a relatórios contábeis municipais. O trabalho de Almeida (2004), por sua vez, avalia a viabilidade de uma plataforma de *Data Warehouse* de baixo custo, sem abordar a execução orçamentária municipal. Portanto, entre os estudos técnicos discutidos nesta seção, não foi identificado um artefato que reúna esse recorte orçamentário, essa esfera de governo e essa forma de modelagem.

No plano internacional, Kalampokis, Tambouris e Tarabanis (2011) propõem um *stage model* para Dados Governamentais Abertos, organizado em duas dimensões: complexidade organizacional e tecnológica, e valor agregado para o consumidor de dados. O modelo sugere que a mera publicação de dados brutos corresponde a um estágio inicial e que a integração entre bases e a agregação analítica constituem estágios superiores. Aplicado ao caso brasileiro, o *stage model* explicita que a Lei de Acesso à Informação e os portais de transparência atuais cumprem o estágio inicial, mas não encerram o ciclo. O *Data Warehouse* proposto neste trabalho se posiciona no estágio de integração e agregação analítica.

Fora do Brasil, alguns projetos práticos consolidaram orçamentos públicos em estruturas analíticas reutilizáveis. O mais conhecido é o *Open Spending*, iniciativa global mantida pela *Open Knowledge Foundation*. Ativo desde o início dos anos 2010, o projeto agrega orçamentos de instituições públicas de diversos países sob um esquema pa-

dronizado e oferece visualizações e uma API. A opção por um modelo de dados genérico foi deliberada para acomodar a heterogeneidade dos sistemas orçamentários nacionais e implica perda de especificidade no tratamento do caso brasileiro, considerando que a estrutura tripartite PPA-LDO-LOA, a classificação MCASP por elemento de despesa e a sequência empenho-liquidação-pagamento não têm correspondência direta no esquema. Para um usuário brasileiro, o *Open Spending* é uma referência arquitetural útil, mas exige adaptação substancial para responder às perguntas que importam no contexto contábil nacional.

A literatura internacional sobre o tema introduz, ainda, o conceito de *citizens' budget* como prática complementar ao DW. Bilge (2015) descreve esse instrumento como uma versão simplificada e explicativa do orçamento público, voltada ao cidadão não especialista e cataloga sua adoção em países como os Estados Unidos, o Reino Unido, a África do Sul e as Filipinas. O ponto importante não é o documento simplificado em si, mas o reconhecimento, em escala internacional, de que a inteligibilidade orçamentária é um problema de *design* de informação, e não apenas de publicação. A solução proposta aqui, via DW, se distingue por permitir a interação analítica, em vez da leitura passiva de um relatório resumido.

No âmbito federal, o Painel do Orçamento Federal disponibiliza dados do SIOP desde 2000 e permite aplicar filtros, gerar visualizações e exportar os resultados (Brasil. Ministério do Planejamento e Orçamento, 2024). O Portal da Transparência da Controladoria-Geral da União reúne dados de sistemas federais e os apresenta em painéis, consultas detalhadas, gráficos e arquivos abertos (Brasil. Controladoria-Geral da União, s.d.). Os dois sistemas tratam dados federais. No nível municipal, a oferta permanece distribuída entre portais com diferentes graus de completude e facilidade de uso (MELO; FUCHIGAMI, 2019; TELES; CARVALHO, 2024). Essa diferença de escala impede que as ferramentas federais atendam diretamente à análise da execução orçamentária de um município.

3.3 *Civic Tech* e Vigilância Civil

Iniciativas de *civic tech* brasileiras complementam o quadro institucional. A Operação Serenata de Amor, o caso mais estudado, partiu de uma abordagem oposta à dos órgãos oficiais. Em vez de tentar cobrir tudo, focou em um problema específico, a verificação de despesas da Cota para Exercício da Atividade Parlamentar, e construiu, sobre ele, algoritmos e visualizações que tornaram públicos casos concretos de irregularidade (POZZO, 2020). Análises acadêmicas do caso mostram que a iniciativa combinou três elementos que raramente aparecem juntos em projetos institucionais, o foco temático estreito, a transparência do próprio código (o projeto é totalmente aberto no GitHub) e o engajamento contínuo com jornalismo de dados como canal de amplificação (NOHARA; COLOMBO, 2019). A lição metodológica é relevante para este trabalho, com foco temático bem definido, produz mais impacto cívico do que a generalidade técnica.

Esse tipo de iniciativa costuma ter vida útil ligada à dedicação de voluntários e a financiamentos pontuais. Silva e Melo (2024) analisam empiricamente a apresentação institucional de iniciativas brasileiras de vigilância civil sobre transparência pública e observam uma tensão entre a sustentabilidade financeira das organizações e a credibilidade de sua atuação como fiscalizadoras. Sistemas mantidos institucionalmente são mais duráveis, mas costumam ser conservadores na escolha do recorte e menos propensos a produzir análises críticas de suas próprias fontes. Há aí uma tensão produtiva entre o experimento de *civic tech* e o sistema institucional. O presente trabalho se posiciona em um ponto intermediário, adotando rigor metodológico de DW combinado com recorte temático claro de uma cidade específica, código aberto disponível para replicação e instância pública que dispensa a instalação local.

3.4 Síntese e Posicionamento deste Trabalho

A Tabela 3.1 sintetiza como este trabalho se posiciona em relação aos principais eixos discutidos.

A Tabela 3.1 mostra diferenças verificáveis no artefato. O sistema integra planilhas do Portal da Transparência de Juiz de Fora, o ementário de receitas da STN e

Tabela 3.1: Posicionamento comparativo em relação à literatura revisada.

Dimensão	Portais avaliados na literatura	Sistemas federais analisados	Este trabalho
Esfera de governo	Municípios ou conjuntos de municípios	Governo Federal	Município de Juiz de Fora
Fontes integradas	Dados publicados pelo portal avaliado	Sistemas orçamentários federais	Portal PJF, ementário STN e LOA municipal
Integração receita–despesa	Avaliada como ausente ou parcial	Disponível em consultas próprias	Reunida em um modelo dimensional
Tratamento das fontes	Inconsistências observadas durante a avaliação	Regras internas não descritas nas interfaces	Regras de transformação, testes e falhas documentados no repositório
Apoio à interpretação	Barreiras de uso identificadas	Filtros e visualizações temáticas	Glossário, <i>tooltips</i> e perguntas que orientam a leitura
Recursos analíticos	Dependem dos recursos de cada portal	Consultas sobre a execução federal	Séries históricas, estágios da despesa e concentração de fornecedores
Forma de acesso	Portais públicos analisados pelos autores	Aplicações web públicas	Aplicação web pública e código para reprodução local
Avaliação	Índices e análise dos portais	Não examinada neste capítulo	Comparação de tarefas com o portal municipal

classificações extraídas da LOA municipal. O modelo resultante contém dez dimensões e três tabelas de fatos para receitas e despesas. As regras de transformação ficam no repositório, e testes estruturais e de domínio verificam os dados a cada materialização. No painel, o usuário encontra séries históricas, comparação entre os estágios da despesa, indicadores de concentração de fornecedores, glossário e *tooltips*. O código, as instruções de execução local e uma instância pública permitem examinar a interface e o processo de produção dos dados. A avaliação compara tarefas executadas no painel e no Portal da Transparência municipal, o que restringe os resultados ao caso estudado.

4 Metodologia

Este capítulo descreve a abordagem metodológica que sustenta o trabalho, as etapas executadas e os critérios adotados em cada uma delas. A escolha metodológica está alinhada à natureza do problema, pois é preciso correlacionar a teoria do controle social a um data warehouse na construção de um artefato concreto, avaliado em relação ao problema que se propõe a resolver.

4.1 Abordagem Metodológica: Design Science Research

O trabalho adota a Design Science Research (DSR) como abordagem metodológica. DSR é orientada à construção e avaliação de artefatos que respondem a problemas práticos relevantes, articulando rigor científico e utilidade prática (HEVNER et al., 2004). Diferentemente de pesquisas puramente descritivas ou explicativas, a DSR assume que parte do conhecimento sobre um domínio é produzida pelo ato de projetar e avaliar artefatos nele. Em ciência da computação aplicada, essa caracterização vem ganhando espaço como alternativa metodológica rigorosa para trabalhos cuja entrega central é um sistema, um modelo ou um processo (DRESCH; LACERDA; JÚNIOR, 2015).

O trabalho executa uma única iteração completa de DSR, desde a definição do problema até a avaliação do artefato. Nesse escopo, a abordagem fornece uma estrutura para relacionar o problema de pesquisa, as decisões de projeto, o sistema construído e os critérios usados na avaliação. Sua vantagem está na rastreabilidade entre essas partes, mesmo sem uma segunda iteração completa. Os resultados se referem à versão avaliada no caso de Juiz de Fora e não demonstram a convergência que poderia ser obtida por ciclos posteriores.

Hevner et al. (2004) propõem sete diretrizes para condução de pesquisas em DSR, das quais quatro orientam decisões concretas deste trabalho:

- **Artefato como pesquisa:** o Data Warehouse e o painel analítico são o produto central, documentado no Capítulo 5 em detalhe suficiente para reprodução.
- **Relevância do problema:** a justificativa do trabalho parte de lacunas amplamente reconhecidas na literatura sobre transparência pública brasileira (Capítulos 1 e 2), não de curiosidade técnica isolada.
- **Avaliação:** o artefato é avaliado contra critérios funcionais explícitos, derivados das perguntas de fiscalização apresentadas na introdução (Capítulo 7).
- **Contribuições claras:** o trabalho declara o que oferece de novo em relação ao estado da arte (Capítulo 3), distinguindo recorte municipal, integração receita-despesa, documentação de inconsistências e articulação teórica.

A única iteração de DSR foi organizada em cinco etapas (DRESCH; LACERDA; JÚNIOR, 2015): *consciência do problema*, *sugestão*, *desenvolvimento*, *avaliação* e *conclusão*. Durante o desenvolvimento, falhas de carga e mudanças nas fontes de dados levaram à revisão de decisões técnicas. Esses ajustes ocorreram na mesma etapa e não constituem novas iterações completas de DSR. A correspondência entre as etapas e os capítulos é a seguinte:

- Consciência do problema: Capítulos 1 e 2.
- Sugestão: Capítulos 2 e 3.
- Desenvolvimento: Capítulo 5.
- Avaliação: Capítulo 7.
- Conclusão: Capítulo 8.

4.2 Revisão Bibliográfica

Para construir a base deste estudo, a revisão da literatura explora três temas principais que se conectam diretamente ao problema investigado: o funcionamento do orçamento público brasileiro, com seus instrumentos, classificações e fases das despesas; a importância da

transparência governamental aliada ao controle social; e a estruturação de dados por meio da modelagem dimensional proposta por Kimball e Ross (2013). Além disso, trazemos as discussões sobre tecnologias cívicas (*civic tech*) aplicadas aos dados públicos, formando um quarto eixo que, embora mais secundário, é fundamental para fornecer o contexto e o posicionamento corretos para a pesquisa.

Neste trabalho, a bibliografia desempenha três papéis metodológicos fundamentais. Primeiro, ela fornece o vocabulário técnico para descrever o cenário orçamentário, incluindo conceitos e siglas essenciais como PPA, LDO, LOA, estágios da despesa, classificação funcional e MCASP. Segundo, ajuda a justificar o problema ao trazer evidências concretas sobre os desafios existentes entre a publicação de dados formais e a prestação de contas (*accountability*) efetiva. E, terceiro, embasa as decisões na construção do nosso artefato, utilizando a metodologia de Kimball como guia principal. A articulação entre a literatura técnica e a literatura sociopolítica foi realizada para compreender até que ponto o uso de uma arquitetura de dados é um objeto político.

4.3 Pesquisa Documental e Seleção de Fontes

A análise documental tem como base principal o Portal da Transparência da Prefeitura de Juiz de Fora. Essa documentação foi complementada por duas fontes externas, Ementário de Classificação por Natureza da Receita, publicado pela Secretaria do Tesouro Nacional (STN), e a Lei Orçamentária Anual publicada pela própria prefeitura.

A seleção das fontes seguiu três critérios. Primeiro, cobertura do ciclo orçamentário completo, com a previsão e a arrecadação da receita, e o empenho, a liquidação e o pagamento da despesa. Segundo, disponibilidade em formato minimamente processável por um pipeline automatizado (planilhas do Excel, mesmo heterogêneas, em vez de PDFs sem estrutura tabular). Quando o formato tabular não estava disponível, como no caso da classificação funcional da LOA, recorreremos à OCR sobre PDFs escaneados. Terceiro, profundidade temporal suficiente para permitir análises de séries históricas.

A pesquisa documental incluiu, ainda, o mapeamento e catálogo dos campos disponíveis em cada fonte, com documentação explícita das heterogeneidades encontradas. As inconsistências catalogadas nessa etapa não foram tratadas como obstáculo a ser

contornado, mas como achado de pesquisa, conforme detalhado na Seção 4.5.

4.4 Desenvolvimento Iterativo do Artefato

A etapa de modelagem seguiu os quatro passos propostos por Kimball e Ross (2013): selecionar os processos de negócio, que, no nosso caso, foram a execução de receitas e despesas, definir o nível de granularidade, identificar as dimensões e estabelecer os fatos. Na prática, esse processo não foi linear. Durante os testes com cargas parciais de dados, precisamos revisar o grão e as dimensões antes de avaliar a versão final do artefato. O desenvolvimento ocorreu em ciclos dinâmicos que alternavam entre programar, carregar amostras de dados e verificar os resultados. Como cada teste revelava novas restrições técnicas, fomos obrigados a revisitar e adaptar decisões anteriores constantemente. Por exemplo, uma mudança inesperada na posição do cabeçalho exigiu a criação de um detector dinâmico de texto; a quebra do *layout* da fonte de dados em maio de 2025 nos forçou a desenvolver dois *parsers* distintos para ler as informações; e a extração da classificação funcional da LOA demandou o uso de reconhecimento ótico de caracteres (OCR). Todos esses ajustes finos fizeram parte da fase de desenvolvimento, compondo a nossa única iteração dentro da metodologia DSR. Para maior clareza, o Capítulo 5 detalha cada um desses obstáculos técnicos e a respectiva regra adotada para solucioná-los.

Esses episódios geraram dois resultados para a pesquisa. No plano técnico, identificamos os pontos de extração que precisam variar, sem alterar o modelo dimensional. No plano conceitual, mostramos que a publicação dos arquivos não garante seu uso em análises históricas, pois mudanças de estrutura e de códigos sem descrição transferem ao usuário a responsabilidade de reconstruir o significado dos dados. A separação entre extração, modelagem e consumo incorpora esse aprendizado à arquitetura. A implementação usa Dagster, Python, DuckDB, dbt e Streamlit e os critérios para essa escolha são discutidos na Seção 4.8.

Além das decisões relativas ao *warehouse* propriamente dito, o desenvolvimento incorporou um princípio de design para a camada de consumo que foi privilegiar uma visão cidadã sobre uma visão puramente contábil. Na prática, isso se traduziu em três escolhas concretas: agrupar códigos oficiais em categorias mais próximas do vocabulário

público, por exemplo, transferências federais e estaduais em lugar de códigos MCASP isolados; acompanhar cada página analítica com um glossário integrado e um bloco orientador com perguntas típicas de fiscalização; e evitar jargão contábil nos títulos e legendas, deslocando-o para *tooltips* acionados por demanda. Essas decisões seguem a leitura freireana desenvolvida no Capítulo 2, na qual a organização da informação não é neutra e condiciona o universo de leitores possíveis.

4.5 Documentação de Inconsistências

A literatura sobre portais municipais aponta a heterogeneidade dos dados publicados como o principal obstáculo à fiscalização (MARCO; TERCI, 2022; MELO; FUCHIGAMI, 2019). Este trabalho busca resolver esse problema, tratando as inconsistências encontradas nas fontes como achados de pesquisa, e não como ruído a ser corrigido silenciosamente.

Em termos práticos, cada anomalia observada durante a extração foi registrada e classificada quanto à natureza, que é mudança de layout, ausência de padronização de cabeçalho, código não resolvido contra o ementário, divergência entre arquivos do mesmo período, e tratamento aplicado, tais como parser alternativo, normalização leve, fallback textual, exclusão documentada. O Capítulo 5 documenta cada uma dessas decisões junto à seção técnica correspondente.

O registro dessas decisões integra a contribuição metodológica do artefato. Para cada inconsistência, a documentação relaciona a característica observada na fonte, o tratamento implementado e a limitação que permanece. Essa relação permite revisar as transformações e indica quais componentes precisam ser adaptados ao aplicar a arquitetura a outro portal municipal.

4.6 Verificação por Invariantes de Domínio

Em dados fiscais publicados, as falhas devem ser tratadas com rigor, pois podem gerar inconsistências. Por exemplo, um valor negativo devido a um erro de parsing, uma quebra de monotonicidade no acumulado anual ou uma dimensão com chave duplicada podem levar a uma conclusão errada sobre a execução orçamentária. A verificação da qualidade

dos dados foi, portanto, incorporada como etapa metodológica explícita, e não como um detalhe técnico.

A verificação opera em duas camadas. A primeira aplica testes estruturais sobre o modelo dimensional para verificar unicidade e não-nulidade de chaves, integridade referencial entre fatos e dimensões, presença das medidas obrigatórias. Esta camada é feita através de declarações em arquivos de configuração do dbt e execução a cada materialização. A segunda aplica invariantes de domínio expressos em SQL, tais como a arrecadação acumulada no exercício que deve ser não decrescente mês a mês, o valor a realizar deve coincidir com a previsão atualizada, menos o arrecadado, e o acumulado anual não pode ser inferior ao acumulado dos meses anteriores na mesma natureza de receita.

A escolha dessas invariantes não é arbitrária: cada uma corresponde a uma classe específica de falha plausível dada a forma como o portal publica os dados (carga parcial entre meses, parsing equivocado, drift de schema). O catálogo completo dos testes está no Capítulo 5.

4.7 Avaliação Funcional Comparativa

A fase de avaliação do ciclo DSR foi implementada como uma análise funcional comparativa entre o Portal da Transparência da Prefeitura de Juiz de Fora e o painel analítico desenvolvido. O critério adotado é funcional e não estético. Assim, dada uma pergunta de fiscalização, quantos passos cada ambiente exige para respondê-la, com que grau de consolidação manual e com que nível de mediação técnica intermediária?

Seleção de tarefas

As tarefas avaliadas foram derivadas das perguntas exemplares levantadas na introdução do trabalho e da literatura sobre uso de portais por organizações de fiscalização (MARCO; TERCI, 2022). O conjunto cobre quatro classes de operação:

- Agregação simples em uma única dimensão (por exemplo, pagamentos totais por função de governo em um exercício).

- Comparação temporal entre exercícios (por exemplo, evolução de investimento em uma área ao longo de três anos).
- Cruzamento entre fontes (por exemplo, previsão versus arrecadação por natureza de receita).
- Análise de concentração (por exemplo, participação dos cinco maiores fornecedores no total pago).

A escolha das classes assegura que a avaliação não se limite a operações em que qualquer ferramenta analítica trivialmente supera uma planilha bruta.

Critérios de comparação

Cada tarefa é avaliada em três dimensões:

- **Número de passos manuais** necessários no portal versus no painel.
- **Necessidade de consolidação** entre arquivos distintos do portal.
- **Grau de mediação técnica** requerido (por exemplo, interpretação de códigos MCASP, decodificação de funcional programática).

O resultado consolidado da comparação está na Tabela 7.1 do Capítulo 7.

Limitações da avaliação

A avaliação foi conduzida pelo autor deste trabalho. Assim, não foram realizados testes de usabilidade com usuários finais, e a comparação foi conduzida pelo próprio autor com base na reprodução repetida das tarefas em ambos os ambientes. Essa limitação é reconhecida explicitamente, sendo que a evidência produzida não diz respeito à compreensão efetiva dos dados por cidadãos sem formação especializada.

A opção pela avaliação autoral tem três justificativas dentro do escopo deste trabalho. Primeiro, o protocolo de comparação é reprodutível, sendo que as tarefas e os critérios estão documentados e podem ser reaplicados por terceiros. Segundo, as tarefas escolhidas estão ancoradas na literatura sobre fiscalização cidadã, e não em preferência

subjetiva. Terceiro, um estudo com usuários exigiria submissão a comitê de ética e recrutamento estruturado, fora do escopo de tempo de um TCC. Essa extensão fica registrada como trabalho futuro no Capítulo 8.

4.8 Compromisso com Reprodutibilidade

Reprodutibilidade foi assumida como princípio metodológico desde o início do projeto, e não como entrega adicional. Para um trabalho que se posiciona como referência metodológica replicável para outros municípios, reprodutibilidade não é opcional, mas sim parte da contribuição.

O compromisso se operacionaliza em quatro decisões concretas. Primeiro, todo o código está versionado em repositório público, com instruções de execução e descrição da pilha tecnológica. Segundo, o pipeline foi projetado para ser idempotente, ou seja, reexecutar uma partição produz o mesmo estado final do warehouse, sem duplicatas nem lacunas. Terceiro, as chaves substitutas das dimensões são geradas por hash determinístico sobre campos naturais, de modo que o reprocessamento preserva identidade entre execuções. Quarto, o painel analítico está publicado em instância pública (*Streamlit Community Cloud*), permitindo que qualquer leitor verifique os resultados sem instalar o ambiente localmente. Os mecanismos técnicos correspondentes estão detalhados no Capítulo 5.

4.9 Cenário de Uso: Juiz de Fora

Juiz de Fora é o cenário no qual a arquitetura foi instanciada e avaliada. Os componentes de extração leem os formatos próprios das fontes municipais, enquanto as camadas de *staging*, modelagem dimensional e consumo mantêm responsabilidades separadas. Essa divisão permite identificar o que pertence ao modelo geral e o que depende do portal escolhido.

O município foi selecionado pela vinculação do trabalho com a UFJF e pela existência de um portal ativo com arquivos que cobrem receitas e despesas em vários exercícios. A aplicação combina as planilhas municipais com o ementário de receitas da STN e com as classificações funcionais extraídas da LOA. O caso também expôs mudanças

de *layout*, cabeçalhos sem posição fixa e documentos que exigiram OCR, condições usadas para avaliar os limites dos componentes de extração.

Os resultados obtidos descrevem as fontes e o período analisados em Juiz de Fora. A aplicação em outro município exige verificar os campos disponíveis e implementar os adaptadores de extração correspondentes. Por esse motivo, o cenário avalia a separação entre os módulos da arquitetura, sem sustentar representatividade estatística para outros portais municipais.

5 Modelagem e Implementação do Data Warehouse

Este capítulo documenta as decisões técnicas na construção do DW proposto, correspondendo à fase de desenvolvimento do ciclo DSR adotado (Capítulo 4). A exposição segue o fluxo dos dados: fontes públicas, extração, *staging*, modelagem dimensional e camada analítica que alimenta os *dashboards*.

5.1 Visão Geral da Arquitetura

A Figura 5.1 apresenta uma visão conceitual da arquitetura implementada. Ela descreve os componentes e seus pontos de extensão, mas ainda não constitui uma arquitetura de referência validada em diferentes organizações. A implementação em Juiz de Fora é a única instância avaliada neste trabalho.

Cada fonte possui um extrator e um tradutor compatíveis com seu formato. Os extratores são *assets* do Dagster que obtêm os arquivos, executam os *parsers* e gravam dados brutos. As *views* de *staging* do dbt traduzem nomes, tipos e códigos para as convenções usadas pelo modelo dimensional. O Dagster registra as dependências e controla a execução por partição. Assim, os componentes específicos de uma fonte ficam antes do *warehouse*.

O acoplamento de uma nova fonte exige um *asset* de extração, um *parser*, a declaração da tabela bruta e um modelo de *staging*. A declaração em `_sources.yml` associa a tabela ao *asset* do Dagster por meio de `meta.dagster.asset_key`. O extrator também deve definir a partição e a estratégia de escrita, além de registrar `nm_arquivo` e `dt_atualizacao`. Se a fonte fornecer conceitos já presentes no *warehouse*, o tradutor produz os campos esperados pelas dimensões e pelos fatos existentes. Um novo processo de negócio exige modelos dimensionais e consultas de consumo próprios. A mesma sequência pode ser usada para dados de outra prefeitura ou de uma instituição como a UFJF, desde

que sejam criados adaptadores para os formatos publicados.

Quatro atributos orientam as decisões descritas neste capítulo. A modularidade delimita os componentes que mudam com cada fonte. A idempotência permite repetir cargas e *backfills*. A testabilidade aparece nos testes estruturais e nas invariantes de domínio. A rastreabilidade liga arquivos de origem, partições, transformações e modelos versionados. Desempenho em grande escala e alta disponibilidade não foram avaliados, pois a implementação usa DuckDB local e uma instância do Streamlit Community Cloud. A testabilidade também tem um limite: os modelos dbt possuem verificações automatizadas, mas os extratores e *parsers* em Python ainda não dispõem de testes unitários.

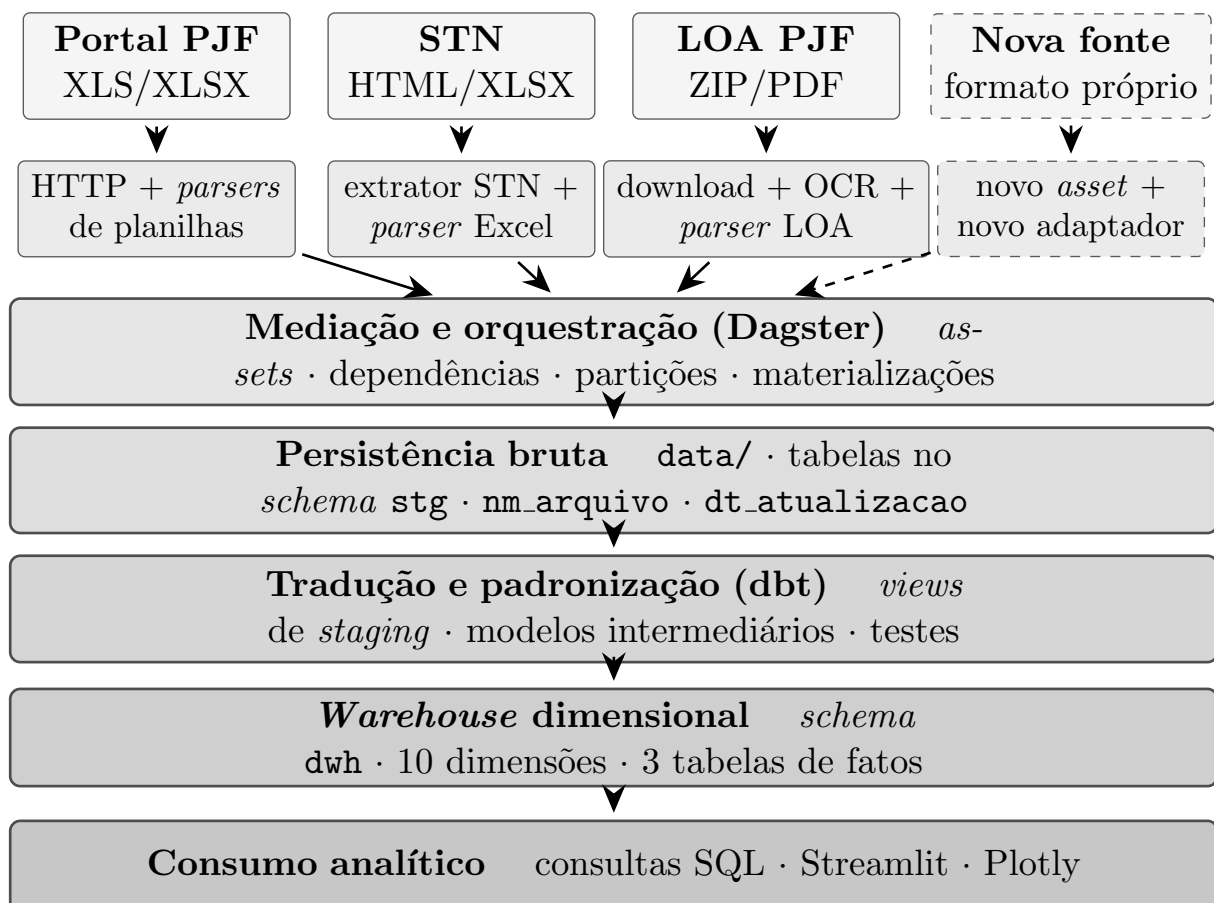


Figura 5.1: Visão conceitual da arquitetura implementada. Os blocos tracejados indicam o ponto de extensão para uma nova fonte; os demais blocos mostram a instanciação usada no cenário de Juiz de Fora.

A arquitetura segue de perto o padrão *medallion* (*bronze/silver/gold*) com adaptações. Os arquivos brutos ficam em `data/`, fora do banco. Tabelas brutas e *views* de *staging* compartilham o *schema stg*: Dagster escreve as tabelas e dbt gera as views. O *warehouse* dimensional ocupa o *schema dwh*. A divisão por *schema* torna explícita a fronteira entre

dado bruto e dado modelado. Quem inspeciona o banco sabe imediatamente o que veio do portal e o que passou por transformações. Alterar um modelo do dbt não obriga a baixar novamente as planilhas do portal.

Todo o código que sustenta as decisões discutidas neste capítulo está versionado em um repositório público (<https://github.com/antoniomarcossouza/UFJF-DCC-TCC>), conforme o compromisso de reprodutibilidade firmado no Capítulo 4.

5.2 Tecnologias Utilizadas

A *stack* foi montada para rodar em máquina local e admitir *deploy* simples, reprodutibilidade local e publicação no Streamlit Community Cloud, com ferramentas já difundidas em engenharia de dados. Para um trabalho com tempo limitado e uma pessoa na implementação, reprodutibilidade foi considerada uma característica importante: quem for replicar o trabalho precisa subir o ambiente sem contratar infraestrutura dedicada. A Tabela 5.1 resume os componentes.

Tabela 5.1: *Stack* técnica utilizada no DW

Componente	Versão	Função no projeto
Python	3.13+	Linguagem de extração e código geral
Dagster	1.12	Orquestrador de <i>assets</i>
pandas	3.0	Leitura e normalização de planilhas
DuckDB	1.5	Banco analítico embarcado, formato OLAP colunar
dbt-core	1.11	Modelagem dimensional em SQL
dbt-duckdb	1.10	Adaptador dbt para DuckDB
dbt_utils	1.3	Macros para chaves substitutas e <i>date spine</i>
Streamlit	1.57	Construção dos <i>dashboards</i>
Plotly	6.7	Visualizações interativas

DuckDB se trata de um banco analítico em processo, sem servidor, com armazenamento colunar voltado a agregações. Para escopo municipal e volume na casa de milhões de linhas, a escolha se apoia em fatores práticos; não há servidor para instalar, o banco é o arquivo `data/execucao_orcamentaria.duckdb`. Consultas analíticas competem, neste porte, com opções pagas. A ponte com **pandas** e Arrow encurta o caminho entre Python e SQL, o que importa quando a extração já roda em **pandas** e a modelagem, em dbt. Como limitação, destaca-se a ausência de suporte robusto à concorrência multiusuário em ambientes de produção. Enquanto o consumo permanecesse restrito a uso local ou

a poucos analistas, essa característica tendia a ter impacto limitado; com a publicação no Streamlit Community Cloud, o volume simultâneo de leitores passa a depender dos limites da plataforma e do arquivo DuckDB empacotado na aplicação.

Dagster e dbt dividem o trabalho de forma comum na engenharia de dados atual: o primeiro orquestra *assets* de forma declarativa, o segundo concentra transformações em SQL. A integração via `dagster-dbt` amarra as duas partes. O *asset* `dbt_non_partitioned_models` só roda depois dos *assets* de extração, de modo que modelagem e carga chegam juntas a cada ciclo. Na prática, quando uma planilha quebra, reexecuta-se Dagster sem tocar no SQL do dbt; quando alguma transformação muda, alteram-se modelos sem reescrever o download.

5.3 Extração: Portal da Transparência, STN e LOA PJF

O Portal da Transparência da Prefeitura de Juiz de Fora não expõe API REST. A coleta depende de HTTP GET sobre planilhas Excel em URLs previsíveis. Esse desenho é típico de portais municipais brasileiros: a transparência existe, mas no formato que a prefeitura já usa internamente (relatórios exportados do SIAFEM), não no formato que um pipeline automatizado preferiria. O município cumpre a Lei de Acesso à Informação; quem quer analisar em série histórica paga o custo de interpretar planilhas heterogêneas. Daí o padrão adotado: baixar arquivo, localizar cabeçalho, interpretar com `pandas`. A Tabela 5.2 sintetiza as fontes consumidas.

As seguintes decisões de extração mostram como o pipeline reage a problemas reais de fonte pública, não a um *dataset* estável de laboratório. Cada uma nasceu de uma falha durante a implementação.

5.3.1 Detecção Dinâmica de Cabeçalho

As planilhas de despesa não fixam a linha de cabeçalho. O arquivo abre com metadados (logotipo, título, período) antes da tabela. A extração percorre as primeiras linhas até achar a célula "Unidade Administrativa" e só então define o início dos dados. Sem essa

Tabela 5.2: Fontes de dados consumidas pelo *pipeline* de extração

<i>Asset</i>	<i>Origem</i>	<i>Partição</i>	<i>Conteúdo</i>
<despesa_mensal consolidada>	PJF	Mensal	Empenhos, liquidações e pagamentos do mês
<receita_mensal comparativa>	PJF	Mensal	Previsão e arrecadação acumuladas no mês
<receita_mensal prevista>	PJF	Anual	Previsão mensal de receitas para o exercício
<stn_ementario_natureza_receita>	STN	Anual	Códigos e descrições oficiais de natureza de receita
<pjf_loa_funcao>	LOA PJF	Sem partição	Funções orçamentárias (DimLOA, Funcao.pdf, exercício 2026)
<pjf_loa_subfuncao>	LOA PJF	Sem partição	Subfunções orçamentárias (DimLOA, SubFuncao.pdf, exercício 2026)

varredura, um mês com uma linha a mais de cabeçalho derruba a carga inteira. Pequenas mudanças de *layout* entre meses deixam de quebrar o pipeline.

5.3.2 Quebra de *Layout* em 2025-05

A planilha de receita comparativa mudou de formato no arquivo de maio de 2025 (partição 2505). Em vez de descartar histórico ou forçar um parser único, o código mantém dois leitores: `read_receita_comparativa_pre2505` e `read_receita_comparativa_2505`. A escolha ocorre na extração, pela chave de partição; assim, anos anteriores continuam reprocessáveis sem perda.

5.3.3 Classificação Funcional da LOA (exercício 2026)

A LOA PJF publica a classificação funcional (DimLOA) em `Funcionais.zip`, com um arquivo PDF por tabela (`Funcao.pdf`, `SubFuncao.pdf`). Os PDFs são imagens escaneadas, não texto selecionável. A extração converte cada página com `pdftoppm` (Poppler) e reconhece o texto com Tesseract (pacotes instalados na imagem Docker). O *parser* valida o layout do exercício 2026 (29 funções, 117 subfunções) e grava duas tabelas brutas em `stg`. Cada exercício novo pode exigir recalibrar o reconhecimento.

Os *assets* não são particionados: o exercício 2026 está fixo na constante `LOA_YEAR`.

Cada ano orçamentário traz PDFs com *layout* distinto; generalizar a ingestão exige validar e, se necessário, adaptar o *parser* ano a ano. Por ora, função e subfunção alimentam `dim_funcao` e `dim_subfuncao`, usadas no *dashboard* para rotular cortes por área de governo. Não há dotação autorizada nesta carga.

5.3.4 Idempotência por Estratégia de Escrita

Cada *asset* usa a estratégia de escrita que combina com sua semântica. Despesa e receita comparativa entram por *append*; a deduplicação fica no *staging*, por `nm_arquivo` e `dt_atualizacao`. O ementário da STN usa *overwrite by partition*, substituindo todas as linhas do ano a cada execução. As tabelas da LOA, anuais e canônicas, usam sobrescrita total da tabela (*overwrite*), pois carregam um único exercício por materialização. Os caminhos diferem, mas o efeito desejado é o mesmo: reexecutar o pipeline devolve o mesmo estado final do *warehouse*; um *backfill* de março de 2023 não deixa duplicata nem lacuna no conjunto de dados.

5.4 Camada de Staging

Seis *views* dbt cobrem as seis fontes brutas (despesa, receita comparativa, receita prevista, ementário STN, função LOA, subfunção LOA). Há deduplicação por `nm_arquivo` e `dt_atualizacao` nas planilhas PJF; remoção de TOTAIS GERAIS na receita comparativa; normalização de códigos de natureza de receita (só dígitos, sem sufixos como .0) e *trim* nas tabelas LOA. Regras mais pesadas (por exemplo, *unpivot* da previsão mensal) ficam em modelos *intermediate* efêmeros, entre *staging* e *warehouse*. Manter o *staging* enxuto evita que uma *view* vire depósito de transformações difíceis de testar.

5.5 Modelo Dimensional

A modelagem segue a metodologia de Kimball: processos de negócio, granularidade, dimensões e fatos. Dois processos foram tratados em separado: execução de despesa e execução de receita. Tentar unificá-los num único fato misturaria grãos, dimensões e calendários de publicação distintos. Despesas e receitas no portal PJF saem em arquivos

diferentes, com colunas e periodicidade próprias; forçar um fato único criaria colunas esparsas e consultas analíticas mais frágeis.

5.5.1 Tabelas de Fatos

O modelo conta com três tabelas de fatos, descritas na Tabela 5.3.

Tabela 5.3: Tabelas de fatos do *warehouse*

Tabela	Granularidade	Medidas
<fct_despesa>	Linha do arquivo mensal consolidado (combinação de empenho, unidade, fornecedor, funcional, natureza, fonte e datas)	<vl_empenhado_mes>, <vl_liquidado_mes>, <vl_pago_mes>
<fct_receita>	Natureza de receita por mês de referência	<vl_previsto_mensal>, <vl_arrecadada_mes>
<fct_receita_acumulada>	Natureza de receita por mês de referência	<vl_previsao_inicial_comparativa>, <vl_previsao_atualizada>, <vl_arrecadada_ano>, <vl_a_realizar>

Uma observação sobre a granularidade de `fct_despesa` é que, no modelo ideal, cada linha seria um empenho ou um estágio (empenho, liquidação, pagamento). O portal, porém, publica relatórios mensais já consolidados: cada linha agrega valores por combinação de dimensões. Modelar nesse grão respeita o que a fonte entrega, sem simular atomicidade inexistente.

`fct_receita` e `fct_receita_acumulada` também nascem de semânticas diferentes. O arquivo de previsão mensal traz valor previsto por mês, alimentando `fct_receita`. O comparativo traz acumulados no ano, alimentando `fct_receita_acumulada`. Juntar tudo numa tabela única exigiria misturar medidas de ritmo mensal com medidas de saldo anual.

5.5.2 Dimensões

O modelo inclui dez dimensões, descritas na Tabela 5.4.

`dim_natureza_receita` cruza duas origens. Quando o ementário da STN cobre o código, prevalece a descrição oficial; nos demais casos, entra a descrição publicada pela PJF. Quem consulta o *dashboard* vê nomenclatura nacional padronizada onde ela existe, sem apagar códigos que só aparecem no município.

Tabela 5.4: Dimensões do *warehouse*

Dimensão	Conteúdo principal
<code>dim_tempo</code>	Dia, mês, trimestre, semestre, ano (<i>date spine</i> de 2014 até 1 ano à frente)
<code>dim_unidade_administrativa</code>	Órgão executor da despesa
<code>dim_funcional_pragmatica</code>	Código e descrição da funcional programática
<code>dim_natureza_despesa</code>	Código e descrição da natureza de despesa (execução PJF)
<code>dim_fonte_recurso</code>	Código e descrição da fonte de recurso
<code>dim_fornecedor</code>	CPF/CNPJ e nome do fornecedor
<code>dim_empenho</code>	Número da nota, modalidade, licitação, referência legal, processo e descrição
<code>dim_natureza_receita</code>	Código e descrição da natureza de receita (enriquecida pelo ementário STN)
<code>dim_funcao</code>	Código e descrição de função (LOA PJF, exercício 2026)
<code>dim_subfuncao</code>	Código e descrição de subfunção (LOA PJF, exercício 2026)

`dim_funcao` e `dim_subfuncao` não se ligam diretamente aos fatos: a execução mensal traz a funcional programática composta em `dim_funcional_pragmatica`. O *dashboard* extrai função e subfunção dessa *string* e faz *join* com as dimensões LOA para exibir nomes oficiais do exercício.

5.5.3 Chaves Substitutas

As dimensões usam chaves substitutas geradas por `dbt_utils.generate_surrogate_key` (MD5 sobre concatenação determinística dos campos naturais). Nas dimensões de despesa derivadas do consolidado mensal, a chave nasce no modelo *intermediate* e é reaproveitada no *warehouse*; `dim_tempo` usa a mesma macro sobre a data do dia. Kimball recomenda esse padrão por dois motivos operacionais. A chave não muda se um rótulo de exibição for corrigido na fonte. A junção entre fatos volumosos e dimensões usa um único `varchar(32)`. O hash determinístico garante que reprocessar o pipeline recria as mesmas chaves, preservando idempotência. Sem chave substituta, uma correção de grafia em `dim_fornecedor` quebraria o histórico de junção ou exigiria atualização em cascata nos fatos.

5.5.4 Esquema Estrela

As ligações entre fatos e dimensões formam três tabelas de fatos articuladas por dimensões conformadas, com `dim_tempo` como dimensão central e `dim_natureza_receita` compartilhada entre as duas tabelas de receita. A Figura 5.3 ilustra essas conexões.

Estrela de Despesa <code>fct_despesa</code> $\leftarrow$ <code>dim_tempo</code> (empenho, liquidação) <code>fct_despesa</code> $\leftarrow$ <code>dim_unidade_administrativa</code> <code>fct_despesa</code> $\leftarrow$ <code>dim_fornecedor</code> <code>fct_despesa</code> $\leftarrow$ <code>dim_funcional_pragmatica</code> <code>fct_despesa</code> $\leftarrow$ <code>dim_natureza_despesa</code> <code>fct_despesa</code> $\leftarrow$ <code>dim_fonte_recurso</code> <code>fct_despesa</code> $\leftarrow$ <code>dim_empenho</code>
Estrela de Receita <code>fct_receita</code> $\leftarrow$ <code>dim_tempo</code> , <code>dim_natureza_receita</code>
Estrela de Receita Acumulada <code>fct_receita_acumulada</code> $\leftarrow$ <code>dim_tempo</code> , <code>dim_natureza_receita</code>

Figura 5.2: Esquemas estrela do *warehouse*, compartilhando `dim_tempo`.

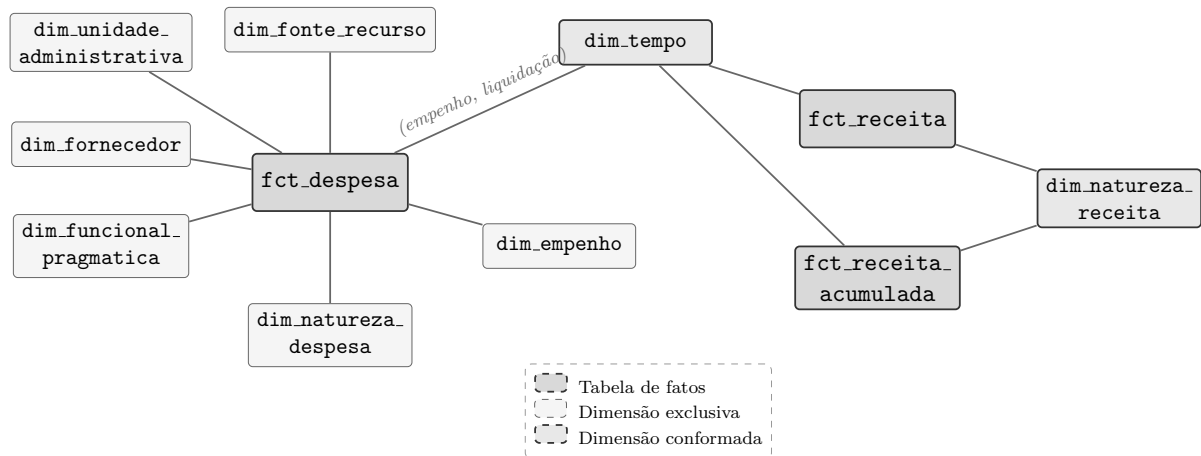


Figura 5.3: Modelo dimensional do *warehouse*: três tabelas de fatos articuladas por dimensões conformadas. `dim_tempo` é compartilhada pelas três fatos e atua em dois papéis (*role-playing*) na estrela de despesa. `dim_natureza_receita` é conformada entre `fct_receita` e `fct_receita_acumulada`.

A dimensão `dim_tempo` aparece na estrela de despesa em dois papéis distintos: data de empenho e data de liquidação. Esse padrão, conhecido como *role-playing dimension* (KIMBALL; ROSS, 2013), evita duplicar a estrutura da dimensão temporal para cada relacionamento. Na prática, a mesma tabela física é referenciada por chaves substitutas distintas nos fatos, permitindo análises que filtram simultaneamente por mês de empenho e por mês de liquidação, sem comprometer a integridade do modelo.

Nenhuma dimensão implementa SCD (*slowly changing dimension*). Todas são tipo 1: guardam só o estado mais recente da fonte. Para descrições de natureza no padrão brasileiro, isso costuma bastar, porque a STN revisa e mentários em ciclos anuais e as mudanças são pontuais. Reorganizações administrativas do município, apesar de raras, ficariam mal registradas: se uma secretaria mudar de nome, o histórico apareceria com o rótulo novo. Capturar essas mudanças exigiria SCD tipo 2.

5.6 Processos de ETL e Idempotência

A unidade de idempotência é a partição temporal nas fontes particionadas: mês de competência nas planilhas PJF e ano de referência na STN. A LOA funcional não usa partição; a idempotência vem da sobrescrita total das tabelas brutas a cada materialização. Reexecutar o pipeline deve devolver o mesmo *warehouse*, independentemente de quantas cargas anteriores existam. Três mecanismos sustentam essa propriedade. A estratégia de escrita por fonte (Seção 5.3) trata STN com sobrescrita por partição, LOA com sobrescrita total e planilhas PJF com *append* mais deduplicação. As chaves substitutas determinísticas evitam linhas falsas em dimensões quando uma partição é recarregada: o mesmo conjunto de campos naturais gera o mesmo MD5. A materialização *table*, e não incremental, nas tabelas de fatos faz o dbt reconstruir o conjunto inteiro a partir do *staging* a cada execução. Para o volume municipal atual, o *full refresh* não compromete desempenho e garante que a tabela final reflita exatamente o que veio do *staging*.

5.7 Qualidade de Dados e Testes

Arquivos `_schema.yml` especificam verificações de unicidade e de nulidade nas chaves das dimensões, de unicidade nas combinações de chaves substitutas nos fatos, de presença das medidas e de integridade referencial. Testes singulares em SQL complementam esse conjunto com invariantes de receita: a arrecadação acumulada no exercício não pode ser inferior ao acumulado dos meses anteriores, e o valor a realizar deve coincidir com a previsão atualizada menos o arrecadado no ano.

Em dados fiscais publicados, falha silenciosa custa caro: um valor negativo por

erro de *parsing* ou uma quebra de monotonicidade entre meses pode virar conclusão errada sobre execução orçamentária. O conjunto de testes não cobre tudo, mas ataca as falhas mais prováveis: chave duplicada em dimensão, medida monetária inválida, carga parcial entre meses e *drift* de *schema* após mudança de *layout* na fonte. A Tabela 5.5 resume os testes singulares.

Tabela 5.5: Testes singulares de qualidade de dados

Teste	Invariante verificada
<dq_fct_receita_acumulada_ao_mes_menor_que_mes>	Valor acumulado no ano deve ser maior ou igual à arrecadação no mês corrente
<dq_fct_receita_acumulada_monotonicidade>	Valor acumulado no ano deve ser não decrescente mês a mês, por natureza
<dq_fct_receita_vl_a_realizar>	Valor a realizar deve coincidir com previsão atualizada menos arrecadado no ano

5.8 Limitações da Implementação

Registrar o que o sistema não faz é tão importante quanto descrever o que faz. Três limitações marcam o recorte atual, delimitando perguntas respondíveis hoje e orientam quem quiser reproduzir ou estender o trabalho.

5.8.1 Granularidade da Despesa

A fonte publica relatórios mensais consolidados, não transações individuais. Análises de empenho a empenho exigiriam outras bases. O modelo define o grão pela linha do arquivo, mas os valores em `fct_despesa` são movimentações mensais por combinação de dimensões, não atos administrativos isolados.

5.8.2 Ausência de Orçamento Autorizado

O *warehouse* traz empenhado, liquidado e pago, mas não a dotação autorizada na LOA. A ingestão atual da LOA cobre apenas a classificação funcional, não os valores autorizados por unidade ou natureza. Sem autorizado, indicadores como percentual de execução sobre o previsto na lei ficam indisponíveis.

5.8.3 Classificação MCASP Funcional Incompleta

A dimensão `dim_funcional_pragmatica` guarda o código composto da funcional programática como *string* única nos fatos. Para rotular função e subfunção no *dashboard*, o sistema consulta `dim_funcao` e `dim_subfuncao`, alimentadas pela LOA PJF do exercício 2026, com *fallback* textual quando o código da execução não consta no ementário carregado.

A revisão de qualidade do *dashboard* quantificou a cobertura dos rótulos LOA na página Execução Orçamentária. Entre os 126 pares distintos função/subfunção presentes na execução, 18 (14,3 %) exibem *fallback* do tipo Função XX ou Subfunção XXX: o código aparece na movimentação mensal, mas não foi resolvido contra `dim_funcao` ou `dim_subfuncao`. Os pares de maior volume executado recebem nomes corretos da LOA; os *fallbacks* concentram-se em combinações de menor pagamento acumulado. Ou seja, o problema afeta sobretudo fatias pequenas da execução, mas ainda compromete a legibilidade de cortes que podem ser relevantes para auditoria.

As seguintes melhorias aparecem como prioridade para reduzir *fallbacks* sem alterar a interface analítica. Revisão manual ou semi-automática dos PDFs após OCR, com tabela de correções versionada (CSV ou *seed dbt*) para códigos sistematicamente omitidos. Enriquecimento cruzado com a tabela MCASP da STN ou com códigos já presentes em `dim_funcional_pragmatica` na execução, usando a LOA PJF como fonte primária e o ementário nacional como reserva para códigos ausentes no PDF municipal. A longo prazo, solicitar à prefeitura publicação dos dados em formato tabular eliminaria a dependência de OCR.

Essas limitações delimitam perguntas respondíveis hoje e apontam extensões concretas: dotação autorizada na LOA, grão mais fino de despesa, desagregação funcional completa, *parser* LOA adaptado a exercícios anteriores e redução dos *fallbacks* funcionais descritos acima.

6 Resultados: Dashboards e Indicadores

O *warehouse* descrito no Capítulo 5 é consumido por cinco páginas analíticas, além das páginas Início e Glossário. Todas leem o banco DuckDB materializado pelo dbt e compartilham filtros na barra lateral. Este capítulo apresenta a interface e os indicadores produzidos pelo artefato construído na única iteração de DSR. A avaliação funcional é tratada separadamente no Capítulo 7.

A organização das páginas segue uma lógica de "visão cidadã": títulos e legendas evitam, quando possível, o uso de jargão contábil; termos técnicos aparecem com explicação no Glossário ou em *tooltips* acionados pelo cursor. Cada página analítica abre um bloco recolhível intitulado "O que você quer descobrir?", que orienta a leitura com perguntas concretas antes de expor gráficos e tabelas. O restante deste capítulo percorre as visualizações principais, descreve os indicadores calculados e contrasta o que o painel permite observar em relação ao Portal da Transparência original.

6.1 Visão Geral da Execução Orçamentária

A leitura consolidada da execução orçamentária distribui-se em duas páginas complementares. A Visão Geral Fiscal responde, em primeiro lugar, se a arrecadação e os pagamentos no recorte filtrado fecham no azul ou no vermelho. Quatro indicadores resumem o recorte: arrecadação, pagamentos, saldo fiscal (diferença entre os dois) e percentual de execução da receita em relação à previsão atualizada. Uma segunda faixa compara a arrecadação e os pagamentos do último mês com os dados disponíveis no filtro de ano. A terceira faixa exibe série mensal de arrecadação *vs.* pagamentos, com eixo vertical em escala logarítmica quando a dispersão de magnitudes exige.

A página Despesas aprofunda os três estágios da despesa pública municipal. No resumo, aparecem valores agregados de empenhado, liquidado e pago, seguidos de razões derivadas dos fatos: percentual liquidado sobre empenhado, percentual pago sobre empenhado e percentual pago sobre liquidado. Esses percentuais não medem execução sobre

dotação autorizada, a informação está ausente na fonte, conforme discutido no Capítulo 5, mas indicam em que medida o compromisso empenhado se converteu em saída efetiva de caixa.



Figura 6.1: Visão Geral Fiscal: resumo do recorte e comparação do último mês (seções 1 e 2).

No portal municipal, receitas e despesas vivem em planilhas distintas, sem cruzamento automático nem série temporal unificada. Empenhado, liquidado e pago aparecem como colunas do relatório mensal, mas comparar estágios ao longo de anos ou filtrar por secretaria exige trabalho manual de consolidação. O painel reduz esse esforço a poucos cliques nos filtros globais e devolve, de imediato, indicadores que antes dependiam de uma planilha montada pelo próprio usuário.

6.2 Execução por Função de Governo

Não foram realizadas entrevistas ou sessões de levantamento de requisitos com servidores da administração municipal. As funcionalidades foram definidas a partir do problema de pesquisa, das perguntas apresentadas na introdução e das barreiras de uso descritas na literatura (MARCO; TERCI, 2022). Por isso, o painel é analisado como instrumento de



Figura 6.2: Visão Geral Fiscal: evolução mensal de arrecadação *vs.* pagamentos (seção 3).

consulta para controle social. Sua adequação ao planejamento de políticas públicas pela administração não foi avaliada.

A página Execução Orçamentária desagrega a despesa por órgão executor e por classificação funcional. A primeira seção ordena as unidades administrativas pelos valores empenhados, liquidados e pagos no período, com um gráfico de barras agrupadas das vinte maiores e uma tabela completa ordenada por pagamento. A segunda seção extrai códigos de função e de subfunção da *string* funcional programática publicada no consolidado mensal e faz *join* com `dim_funcao` e `dim_subfuncao`, alimentadas pela LOA PJF do exercício 2026.

A página permite executar a pergunta usada como exemplo na introdução: *"Quanto Juiz de Fora investiu em educação nos últimos três anos?"*. O usuário seleciona os exercícios e consulta as linhas classificadas na função Educação. Essa tarefa representa uma comparação temporal por função de governo e foi escolhida para verificar a capacidade analítica do painel. Sua prioridade para contribuintes ou gestores não foi validada com esses grupos.

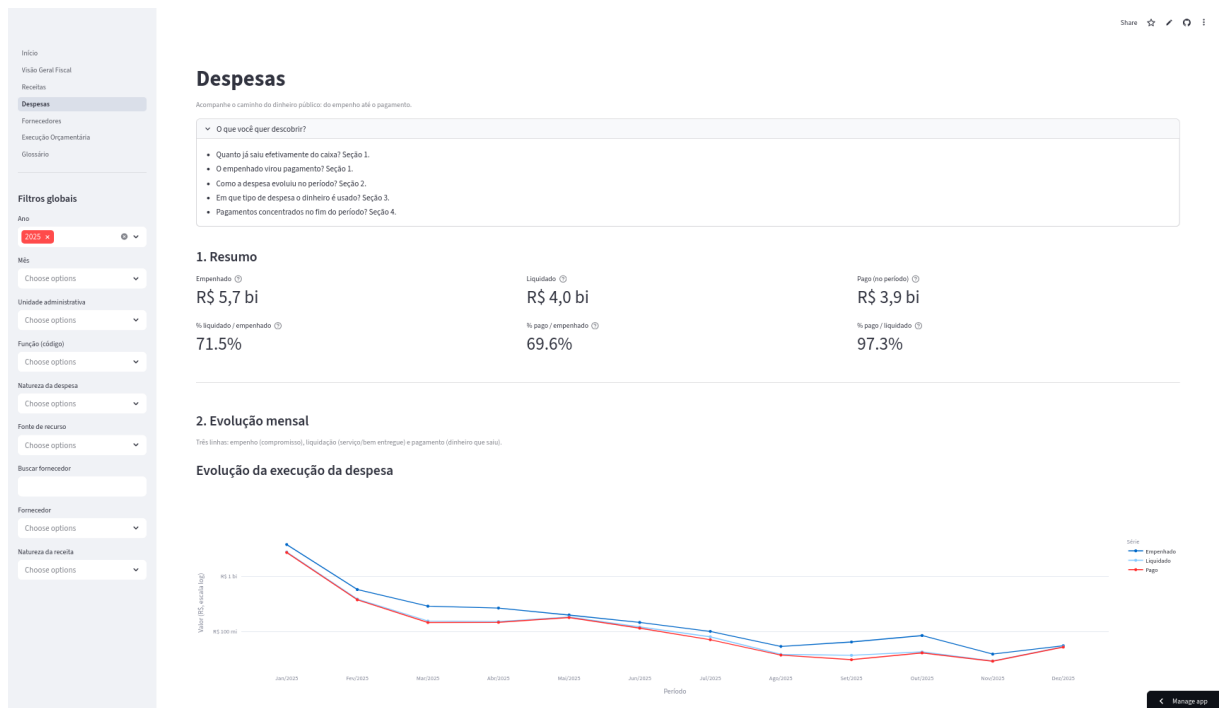


Figura 6.3: Despesas: resumo dos estágios da execução e da evolução mensal empenhada, liquidada e paga.

6.3 Concentração de Fornecedores

A página Fornecedores trata da questão do favorecimento e da concentração de recursos. O resumo destaca o maior favorecido do ranking, a base total de pagamentos considerada e a parcela acumulada pelos cinco primeiros do ranking. A seção de concentração exibe um gráfico horizontal com os vinte maiores pagamentos e uma tabela com valor pago, quantidade de empenhos distintos e participação percentual no total do recorte. A terceira seção permite escolher um fornecedor e visualizar a evolução mensal dos pagamentos a ele, o que é útil para verificar se um contrato pontual distorce o ranking ou se há recebimento recorrente ao longo do tempo.

O portal publica o favorecido e o valor pago na planilha de despesa, mas não calcula a participação relativa nem um ranking dinâmico filtrável por área ou período. Identificar concentração exige exportar, ordenar e normalizar manualmente, tarefa factível para, por exemplo, um jornalista, porém improvável para o cidadão sem intermediário. O indicador “% nos 5 maiores” condensa essa análise em um número legível e convida à pergunta de fiscalização: a dependência de poucos credores é compatível com a natureza dos serviços contratados?



Figura 6.4: Execução Orçamentária por função e subfunção: gráfico das vinte maiores combinações e tabela funcional.

6.4 Receitas Municipais

A página Receitas organiza a entrada de recursos. O resumo contrasta previsão atualizada, arrecadação acumulada no recorte, razão realizado/previsão e parcelas classificadas como repasse federal e estadual, utilizando como base prefixos de natureza de receita. A seção "De onde vem o dinheiro" agrupa os códigos MCASP por origem interpretável e exibe barras horizontais por origem. Duas séries temporais comparam previsto e realizado: uma anual (ignorando o filtro de ano, mantendo mês e natureza) e outra mensal. A seção final "Previsto *vs.* realizado" usa uma barra de progresso agregada e uma tabela por natureza de receita, com percentual de realização linha a linha.

Cruzar previsão e arrecadação no portal implica reconciliar arquivos comparativos e de previsão mensal, além de interpretar códigos de natureza sem ementário unificado. O *warehouse* já integra descrições da STN quando disponíveis; o painel traduz a composição da receita em categorias próximas do debate público, o que aproxima o dado da pergunta "de onde vem o dinheiro da cidade?".



Figura 6.5: Fornecedores: concentração de pagamentos (top 20) e ranking completo com participação percentual.

6.5 Análise dos Dados de Juiz de Fora

Descrever painéis não encerra o capítulo de resultados. A proposta do trabalho, ancorada na tradição freireana discutida no Capítulo 2, exige ir além da descrição da interface e verificar o que o conjunto de números sugere ao ser lido pelo próprio artefato construído. Três padrões emergem da exploração sistemática do *dashboard* sobre a base materializada.

Dependência de transferências. A composição por origem mostra, de forma recorrente, que transferências correntes respondem por parcela expressiva da arrecadação municipal. Isso não é específico de Juiz de Fora, mas torna visível uma premissa muitas vezes invisível no debate local: parte relevante do “orçamento da cidade” depende de repasses cujas regras são definidas fora do município. Um conselheiro ou vereador pode usar o painel para calibrar expectativas sobre o que a prefeitura controla diretamente *vs.* o que é efeito de política nacional ou estadual.

Distância entre empenho e pagamento. Nas páginas de Despesas e Execução Orçamentária, o empenhado tende a superar o liquidado e pago no mesmo recorte quando o período inclui meses recentes do exercício. Isso reflete o ciclo contábil normal, mas também alerta contra leituras simplistas: “gastou” no sentido de empenho não coincide

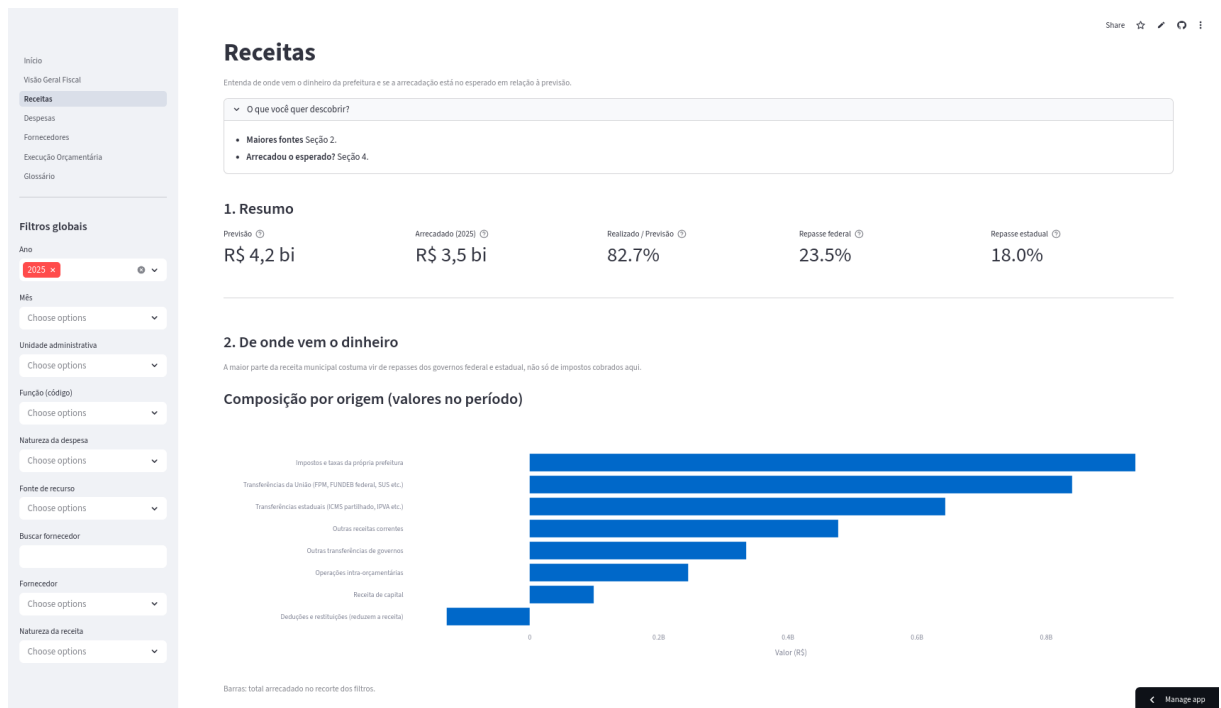


Figura 6.6: Receitas: resumo do recorte e composição por origem.

com "saiu do caixa". Restos a pagar e atrasos de liquidação só seriam mensuráveis com granularidade transacional ou com dotação autorizada, limitações já discutidas.

Legibilidade funcional incompleta. Os *fallbacks* de rótulo LOA (Capítulo 5) concentram-se em combinações de menor volume, mas ainda aparecem em cortes que um fiscalizador pode escolher deliberadamente. A análise cidadã ganha densidade quando o leitor combina a tabela funcional com o Glossário e, diante de **Função XX**, consulta o código ou busca esclarecimento junto ao órgão público.

No painel público, o cidadão formula a pergunta antes de interagir com a interface e procura respondê-la por meio das páginas e dos filtros disponíveis. Não há campo de texto livre nem uma área para executar SQL. O conjunto de consultas é definido pelo código da aplicação, com parâmetros restritos aos filtros. Essa escolha simplifica o uso e impede consultas arbitrárias, mas limita perguntas que não tenham sido previstas no desenho das páginas.

Usuários com conhecimento técnico podem reproduzir o *pipeline*, abrir o banco DuckDB localmente em modo somente leitura e escrever consultas próprias. Disponibilizar esse recurso na aplicação pública exigiria uma camada isolada de consulta, com validação das instruções, bloqueio de acesso ao sistema de arquivos e limites de tempo e volume

de resposta. Uma conexão somente leitura, isoladamente, não controla abuso de recursos nem todas as operações permitidas pelo banco. A consulta SQL pública fica, portanto, registrada como possível extensão, e não como funcionalidade do artefato avaliado.

Os três padrões descritos acima ilustram o tipo de leitura que o painel viabiliza, mas não esgotam tudo o que ele permite. O Capítulo 7 compara, de forma estruturada, o esforço necessário para realizar essas leituras no painel e no portal original.

7 Avaliação e Discussão

Este capítulo verifica o artefato em relação ao problema de pesquisa e aos objetivos definidos no Capítulo 1. A avaliação compara as operações necessárias para executar as mesmas tarefas no Portal da Transparência e no painel. O procedimento foi realizado pelo autor e não incluiu participantes.

7.1 Avaliação Funcional

A ISO 9241-11 define usabilidade em relação a usuários, objetivos e contextos específicos, considerando eficácia, eficiência e satisfação (International Organization for Standardization, 2018). Experiência do usuário abrange ainda aspectos que não se limitam ao desempenho em tarefas (HASSENZAHN; TRACTINSKY, 2006). Como este trabalho não observou usuários nem mediu satisfação, tempo ou aprendizado, seus resultados não constituem uma avaliação de usabilidade ou de experiência do usuário.

O critério adotado é funcional. Para cada pergunta de fiscalização, registram-se as operações exigidas, a necessidade de consolidar arquivos e o conhecimento técnico necessário nos dois ambientes. A Tabela 7.1 apresenta as seis tarefas comparadas. Um estudo posterior deve observar participantes em contextos definidos, medir a conclusão das tarefas e investigar o que eles aprendem durante o uso.

Nas tarefas da tabela, o portal fornece os dados, mas exige que o usuário integre arquivos ou interprete códigos fora da interface. O painel reutiliza as transformações materializadas no *warehouse*. A comparação demonstra redução das operações manuais nesses seis casos; ela não permite concluir como usuários reais perceberiam a interface.

7.2 Limitações do Trabalho

As limitações do *warehouse* (Capítulo 5) propagam-se ao *dashboard*. Ausência de dotação autorizada impede indicadores de execução sobre o previsto na LOA, métrica padrão em

Tabela 7.1: Comparação portal PJF *versus dashboard* em tarefas de fiscalização

Pergunta	Portal da Transparência	<i>Dashboard</i>
Saldo entre arrecadação e pagamentos no ano	Baixar planilhas de receita e despesa; alinhar competência; somar colunas manualmente	Página Visão Geral Fiscal; filtrar ano
Evolução mensal empenhado/liquidado/pago	Vários arquivos mensais; colar e consolidar	Página Despesas; gráfico único
Pagamentos por função	Filtrar planilha por código funcional; somar; decodificar código	Página Execução Orçamentária; filtrar anos; ler tabela com rótulos LOA
Participação dos maiores fornecedores	Exportar; ordenar; calcular % no Excel	Página Fornecedores; indicador "% nos 5 maiores"
Previsão <i>vs.</i> arrecadação por natureza	Reconciliar comparativo e previsão; interpretar códigos	Página Receitas; tabela com % de realização
Significado de "liquidação"	Buscar glossário externo ou norma	Glossário integrado + <i>tool-tips</i>

relatórios de controle interno. O grão por empenho existe no *warehouse*, e `dim.empenho` guarda a descrição e os metadados de licitação. O que falta é o detalhe abaixo do empenho. O pagamento aparece apenas como total mensal, sem data própria, o que dificulta a medição de prazos até a quitação. Itens, quantidades e preços de cada empenho não são publicados, e a interface ainda não usa esses metadados. Rótulos LOA com *fallback* em cerca de 14% dos pares função/subfunção presentes na execução comprometem a legibilidade pontual.

Limitações próprias da implementação do painel acrescentam-se às da fonte. O painel está publicado no *Streamlit Community Cloud*¹, lendo um DuckDB em arquivo empacotado na aplicação. O orquestrador Dagster, porém, não roda de forma agendada e, por isso, a atualização exige reexecutar manualmente o *pipeline* e republicar o banco de dados. Os dados exibidos refletem, portanto, a última materialização, não necessariamente o portal em tempo real. Por fim, o cruzamento com licitações e contratos, citado na introdução como uma lacuna dos portais, permanece fora do escopo, embora conste como trabalho futuro.

¹<https://juiz-de-fora-transparencia.streamlit.app/>

7.3 Implicações para Transparência e Controle Social

Retomando as três lacunas da introdução, o sistema construído endereça cada uma delas em graus distintos.

A lacuna técnica é a mais diretamente atacada. Integrar receitas e despesas, conciliar planilhas heterogêneas, normalizar naturezas com o ementário STN e expor séries temporais filtráveis eliminam o gargalo de consolidação manual que a literatura associa aos portais municipais (MARCO; TERCI, 2022; MELO; FUCHIGAMI, 2019). O cidadão deixa de depender de uma planilha intermediária montada por um especialista ao acessar o *dashboard* em uma URL pública, sem instalar o ambiente localmente.

A lacuna social é parcialmente mitigada. Glossário, linguagem de "visão cidadã" e agrupamento das naturezas de receita por origem aproximam os termos do MCASP do vocabulário público. Permanecem os códigos nos filtros e a distância entre a classificação contábil e a experiência vivida: o painel mostra quanto foi pago em "saúde" na classificação funcional, não quanto foi investido em um equipamento específico no bairro do usuário. Transparência agregada educa sobre prioridades macro; não substitui prestação de contas situada.

O aspecto pedagógico foi traduzido em decisões de interface: perguntas orientadoras, glossário, *tooltips* e comparações temporais. A referência a Freire (FREIRE, 1967) orienta a intenção de favorecer a formulação e a interpretação de perguntas. A modelagem de Kimball (KIMBALL; ROSS, 2013) fornece a estrutura usada para calcular e apresentar as respostas. O painel, porém, oferece consultas predefinidas e filtros; ele não recebe perguntas livres. Também não foram coletadas evidências de aprendizagem ou de mudança na capacidade de fiscalização dos usuários. Assim, o resultado avaliado é a disponibilidade de recursos de apoio à leitura, e não a ocorrência de conscientização.

7.4 Reprodutibilidade, Replicabilidade e Operação

O repositório público contém o código da extração, os modelos dbt, a aplicação Streamlit, o arquivo de versões das dependências e a configuração Docker. As instruções para instalar o ambiente, executar o *pipeline* e iniciar o painel estão no arquivo README do

repositório².

A replicação para outro município não foi testada. A separação entre extração, *staging*, modelo dimensional e consumo permite reutilizar parte da estrutura, mas cada portal requer levantamento de campos, *parsers*, regras de escrita e testes próprios. A arquitetura oferece pontos de extensão; não há evidência empírica de que a adaptação seja simples ou tenha o mesmo resultado em outras cidades.

A operação contínua também permanece incompleta. O painel de Juiz de Fora está publicado, mas a atualização do banco e sua republicação são manuais. Manter o serviço requer agendamento, monitoramento das fontes e definição de um responsável. Uma universidade ou organização de controle social poderia assumir essa função, mas não há acordo institucional estabelecido. Os próximos passos são automatizar a atualização e realizar uma avaliação com usuários nos contextos de uso definidos.

²<https://github.com/antoniomarcossouza/UFJF-DCC-TCC>

8 Conclusão

8.1 Síntese das Contribuições

A pergunta de pesquisa foi respondida pela organização dos dados do Portal da Transparência de Juiz de Fora em um modelo dimensional alimentado por um *pipeline* reexecutável. O modelo integra receitas e despesas, preserva as classificações orçamentárias e fornece descrições e indicadores usados pelo painel. Na comparação funcional, essa estrutura reduziu as operações manuais necessárias nas seis tarefas avaliadas. O resultado se restringe às barreiras técnicas e ao caso estudado: não foram realizados testes com usuários nem a adaptação para outro município.

O primeiro objetivo previa a coleta e documentação de dados de receitas e despesas do Portal da Transparência da Prefeitura. Foi atendido: o *pipeline* de extração consome seis fontes (despesa mensal consolidada, receita mensal comparativa, receita mensal prevista, ementário de natureza de receita da STN, função LOA e subfunção LOA do exercício 2026), com tratamento explícito de heterogeneidades como mudança de *layout* no meio do histórico, cabeçalho de planilha sem posição fixa e OCR nos PDFs da LOA.

O segundo objetivo, que previa a modelagem dimensional conforme o modelo de Kimball, foi cumprido. O *warehouse* conta com dez dimensões e três tabelas de fatos organizadas em esquema de estrela compartilhando `dim_tempo`. Cada decisão de granularidade foi escolhida em função do que a fonte de fato publica, sem inventar atomicidade que não está presente nos dados originais.

O terceiro objetivo previa a implementação de processos de ETL e foi atendido com Dagster como orquestrador, Python e *pandas* para extração, DuckDB como banco de dados analítico e dbt para modelagem em SQL. A unidade de idempotência foi definida como uma partição temporal, e três mecanismos garantem que a reexecução dessa mesma partição produza o mesmo estado final do *warehouse*.

O quarto objetivo, que previa a construção de *dashboards* interativos, também foi alcançado. Sete páginas em Streamlit consomem o *warehouse*, estão publicadas em

uma URL pública e respondem perguntas como a execução orçamentária por área, a evolução temporal de receitas, a concentração de fornecedores e o desvio entre previsão e arrecadação. A exploração sistemática descrita no Capítulo 6 destacou duas leituras recorrentes: dependência de transferências intergovernamentais na arrecadação e a distância entre empenho e pagamento ao longo do exercício.

Complementarmente, o Capítulo 7 compara o portal e o *dashboard* em seis tarefas de fiscalização. O painel reduz os passos manuais de consolidação em todas elas. A lacuna técnica dos portais foi a mais atendida, e a pedagógica, parcialmente.

8.2 Considerações Finais

O processo de construção deste *warehouse* revelou algo que não estava previsto no plano inicial. O problema central da transparência orçamentária municipal não é técnico no sentido mais estreito da palavra.

O problema está na distância entre o que o portal publica e o que o cidadão precisa para fiscalizar. O portal publica relatórios mensais consolidados em planilhas Excel. O cidadão precisa de séries temporais comparáveis, classificações inteligíveis e indicadores que dialoguem com sua experiência cotidiana. Reduzir essa distância é trabalho de arquitetura de dados, e arquitetura de dados é, em última análise, uma decisão sobre quem deve interpretar a informação pública. Quando se publica o valor apenas em colunas mensais, sem a data de cada pagamento e sem abrir o empenho por itens, decide-se o que o cidadão pode ver. Ele enxerga quanto saiu no mês e a descrição do empenho. Não enxerga o ritmo dos pagamentos, nem a quantidade nem o preço do que foi comprado.

A leitura freireana ganha contorno técnico nesse ponto. Conscientização orçamentária, traduzida para o trabalho de quem constrói *data warehouses*, significa modelar o dado pensando em quem vai lê-lo, e não apenas em quem vai gerá-lo.

8.3 Trabalhos Futuros

A continuação natural do trabalho segue cinco direções concretas:

- (1) Integração da dotação autorizada pela LOA, viabilizando indicadores percen-

tuais de execução sobre o autorizado, que são a métrica padrão utilizada por órgãos de controle. A classificação funcional do exercício de 2026 já está no DW.

(2) Desagregação da funcional programática em programa e ação, com *parsers* LOA adaptados a exercícios anteriores e, se necessário, aos ementários oficiais da STN para níveis ainda ausentes.

(3) Captura de licitações e contratos como processos de negócio adicionais, permitindo rastrear a cadeia completa do gasto e analisar o tempo médio entre as etapas.

(4) Implementação de SCDs do tipo 2 nas dimensões sujeitas a mudanças históricas, permitindo análises "como era" *vs.* "como é".

(5) Replicação do modelo em pelo menos um município de porte semelhante, com análise comparativa, para validar a generalidade do esquema dimensional proposto e identificar pontos de adaptação locais.

Cada uma dessas direções é viável com o ferramental atual e não exige reescrita das camadas já implementadas. O DW aqui construído, quando tomado em conjunto com a discussão teórica que o sustenta, não esgota o problema da transparência orçamentária. Oferece, em vez disso, uma peça de evidência de que é possível, no escopo de um trabalho, trilhar o caminho entre o dado bruto que a prefeitura publica e a pergunta que o cidadão de fato faria. Essa ponte é a contribuição.

Bibliografia

ALMEIDA, E. C. d. *Estudo de viabilidade de uma plataforma de baixo custo para Data Warehouse*. Dissertação (Dissertação (Mestrado em Informática)) — Universidade Federal do Paraná, Curitiba, 2004. Orientador: Prof. Dr. Marcos Sfair Sunye.

BALDO, R. A. Democratização do orçamento público pela legalidade, legitimidade e economicidade. *Revista Brasileira de Políticas Públicas*, Brasília, v. 8, n. 1, p. 689–705, 2018.

BILGE, S. A new approach in public budgeting: Citizens' budget. *Journal of International Education and Leadership*, v. 5, n. 1, 2015. ISSN 2161-7252.

BRASIL. *Lei nº 12.527, de 18 de novembro de 2011*. Brasília, DF, 2011. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm.

BRASIL. *Manual de Contabilidade Aplicada ao Setor Público*. 2024. Disponível em: <https://cnm.org.br/storage/noticias/2024/Links/MCASP%20-%2011%C2%AA%20Edi%C3%A7%C3%A3o.pdf>.

BRASIL. *Manual Técnico de Orçamento – MTO*. 2024. Disponível em: <https://bibliotecadigital.gestao.gov.br/handle/777/234?offset=20>.

Brasil. Controladoria-Geral da União. *O que é e como funciona: Portal da Transparência do Governo Federal*. s.d. Disponível em: <https://portaldatransparencia.gov.br/sobre/o-que-e-e-como-funciona>.

Brasil. Ministério do Planejamento e Orçamento. *Consultas a bases de dados: Painel do Orçamento Federal*. 2024. Disponível em: <https://www.gov.br/planejamento/pt-br/acesso-a-informacao/acoes-e-programas/carta-de-servicos/consultas-a-bases-de-dados>.

CARDOSO, P. H. *Ciência de dados aplicada a dados governamentais abertos sob a ótica da Ciência da Informação*. 110 p. Dissertação (Dissertação (Mestrado em Ciência da Informação)) — Universidade Federal do Rio de Janeiro; Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2019. Orientador: Prof. Dr. Marcelo Fornazin.

DRESCH, A.; LACERDA, D. P.; JÚNIOR, J. A. V. A. *Design Science Research: método de pesquisa para avanço da ciência e tecnologia*. Porto Alegre: Bookman, 2015.

FREIRE, P. *Educação como prática da liberdade*. [S.l.]: Editora Paz e Terra, 1967.

FREIRE, P. *Pedagogia da autonomia: saberes necessários à prática educativa*. [S.l.]: Editora Paz e Terra, 1996.

GIACOMONI, J. *Orçamento Público*. 18. ed. [S.l.]: Atlas, 2021. ISBN 9788597026122.

HASSENZAHN, M.; TRACTINSKY, N. User experience – a research agenda. *Behaviour & Information Technology*, v. 25, n. 2, p. 91–97, 2006.

HEVNER, A. R. et al. Design science in information systems research. *MIS Quarterly*, v. 28, n. 1, p. 75–105, 2004.

International Organization for Standardization. *Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts*. 2018. Disponível em: <https://www.iso.org/standard/63500.html>.

KALAMPOKIS, E.; TAMBOURIS, E.; TARABANIS, K. Open government data: A stage model. In: JANSSEN, M. et al. (Ed.). *Electronic Government (EGOV 2011)*. Berlin, Heidelberg: Springer, 2011, (Lecture Notes in Computer Science, v. 6846). p. 235–246.

KIMBALL, R.; ROSS, M. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3. ed. Indianapolis, IN: John Wiley & Sons, 2013. ISBN 9781118530801.

MARCO, C. A. F. D.; TERCI, E. T. Transparência municipal e controle social: a visão dos observatórios sociais sobre os portais de transparência e acesso à informação. *Interações (Campo Grande)*, SciELO Brasil, v. 23, n. 2, p. 313–330, 2022.

MATIAS-PEREIRA, J. *Curso de administração pública: foco nas instituições e ações governamentais*. 3. ed. São Paulo: Atlas, 2010.

MELO, D. A.; FUCHIGAMI, H. Y. Proposta de índice bidimensional de transparência da informação público-eletrônica como ferramenta para participação e controle sociais. *REAd. Revista Eletrônica de Administração (Porto Alegre)*, SciELO Brasil, v. 25, n. 2, p. 179–214, 2019.

NOHARA, I. P.; COLOMBO, B. A. Tecnologias cívicas na interface entre direito e inteligência artificial: Operação serenata de amor para gostosuras ou travessuras? *A&C – Revista de Direito Administrativo & Constitucional*, Belo Horizonte, v. 19, n. 76, p. 83–103, 2019.

NOVAES, F. S.; SANTOS, M. E. P. d. O orçamento participativo e a democratização da gestão pública municipal – a experiência de vitória da conquista (ba). *Revista de Administração Pública*, SciELO Brasil, v. 48, p. 797–820, 2014.

PINHO, J. A. G. d.; SACRAMENTO, A. R. S. Accountability: já podemos traduzi-la para o português? *Revista de Administração Pública*, SciELO Brasil, v. 43, p. 1343–1368, 2009.

POZZO, F. D. *Operação Serenata de Amor: análise de um projeto de tecnologia cívica que usa inteligência artificial para auditar contas públicas*. Dissertação (Dissertação (Mestrado em Políticas Públicas)) — Universidade Federal do Rio Grande do Sul, Porto Alegre, 2020.

SILVA, D. R.; MELO, F. S. d. A. A dupla transparência nas apresentações de si de iniciativas brasileiras de vigilância civil sobre transparência pública. *Comunicação e Sociedade*, v. 46, p. e024019, 2024.

SOUZA, A. W. d. A.; MIRANDA, M. G. d.; FARIAS, B. M. d. Transparência e controle social no orçamento participativo: desafios e perspectivas para a democracia local no brasil. *Revista Científica Multidisciplinar O Saber (RCMOS)*, V, n. 1, 2025. ISSN 2675-9128.

STRECK, D. R.; PITANO, S. d. C.; MORETTI, C. Z. Educar pela participação, democratizar o poder: o legado freireano na gestão pública. *Educação em Revista*, SciELO Brasil, v. 33, p. e167880, 2017.

TELES, J. A. d. C.; CARVALHO, K. M. d. Panorama da produção acadêmica sobre transparência pública municipal no brasil: uma revisão de escopo dos últimos 35 anos. *Revista da AGU*, v. 23, n. 1, 2024.