



UNIVERSIDADE DO SUL DE SANTA CATARINA

ANDRÉ LUIZ PIZANI

HIROSHI BUYO SIQUEIRA

**GERAÇÃO DE INDICADORES DA JUSTIÇA:
UTILIZANDO UMA SOLUÇÃO DE BUSINESS INTELLIGENCE COM DADOS
ABERTOS**

Florianópolis

2018

ANDRÉ LUIZ PIZANI
HIROSHI BUYO SIQUEIRA

**GERAÇÃO DE INDICADORES DA JUSTIÇA:
UTILIZANDO UMA SOLUÇÃO DE BUSINESS INTELLIGENCE COM DADOS
ABERTOS**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Sistemas de Informação da Universidade do Sul de Santa Catarina, como requisito parcial à obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Flávio Ceci, Dr.

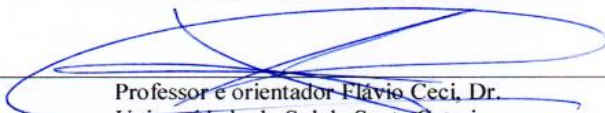
Florianópolis
2018

ANDRÉ LUIZ PIZANI
HIROSHI BUYO SIQUEIRA

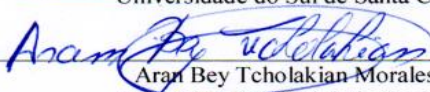
**GERAÇÃO DE INDICADORES DA JUSTIÇA:
UTILIZANDO UMA SOLUÇÃO DE BUSINESS INTELLIGENCE COM DADOS
ABERTOS**

Este Trabalho de Conclusão de Curso foi julgado adequado à obtenção do título de Bacharel em Sistemas de Informação e aprovado em sua forma final pelo Curso de Graduação em Sistemas de Informação da Universidade do Sul de Santa Catarina.

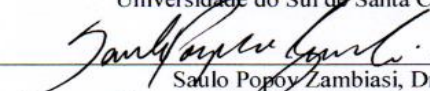
Florianópolis, 19 de novembro de 2018.



Professor e orientador Flávio Ceci, Dr.
Universidade do Sul de Santa Catarina



Aran Bey Tcholakian Morales, Dr.
Universidade do Sul de Santa Catarina



Saulo Popov Zambiasi, Dr.
Universidade do Sul de Santa Catarina

AGRADECIMENTOS

Agradeço em primeiro lugar ao meu pai Marcos, minha mãe Luciana e a minha avó Teresinha, por todo o apoio que me deram nessa jornada e por sempre terem feito de tudo para essa e muitas outras conquistas ser tornassem realidade.

Ao meu grande amigo e dupla deste trabalho Hiroshi, pela dedicação, conselhos e aprendizados que vão além da vida profissional.

Ao professor orientador desta monografia Dr. Flávio Ceci, por ter acreditado e apoiado este trabalho, além de ter me incentivado a trabalhar na área de dados.

Ao meu primo Gilberto, que sempre serviu de grande exemplo e incentivo dentro da área de TI.

Aos demais profissionais com quem trabalhei, professores, amigos e familiares que, de alguma forma, estiveram presentes ao longo desses últimos anos, me proporcionando grandes experiências e aprendizados.

André Luiz Pizani

À minha mãe Noemi e ao meu pai Thomaz, por sempre me apoiarem em todas as situações, pelos conselhos, incentivos e sempre me ensinando com palavras e principalmente por ações a ser uma pessoa melhor.

À minha irmã Joana pela sua amizade, carinho e por sempre me fazer rir nos momentos que mais preciso.

Ao meu caro amigo André pela paciência, honestidade, disponibilidade e empenho neste trabalho, na vida profissional e pessoal.

Agradeço ao professor orientador Dr. Flávio Ceci, por toda sua ajuda e colaboração neste trabalho, pelo seu profissionalismo, didática e companheirismo, que nos ensinou com seu jeito animado a beleza e a importância dos dados nas nossas vidas.

Aos colegas de faculdade, de trabalho, e aos demais professores por proporcionar um ambiente de aprendizado constante, de compartilhamento de conhecimentos e experiências.

Hiroshi Buyo Siqueira

RESUMO

Este trabalho tem como objetivo apresentar e implementar uma solução de *Business Intelligence* (BI) com os dados abertos disponibilizados pelo Conselho Nacional de Justiça (CNJ). Esses dados divulgam a realidade dos tribunais brasileiros a respeito de sua estrutura, litigiosidade e análises essenciais para a gestão judiciária. A referência bibliográfica traz como objetivo conceituar todo o processo de uma solução de BI, desde a estruturação até o *Online Analytical Processing* (OLAP) e sobre o judiciário. A modelagem do sistema foi desenvolvida com base nas perguntas estratégicas levantadas. O sistema é composto pela criação de um *Staging Area*, a criação de um *Data Warehouse* (DW), na criação de processos de *Extract, Transform and Load* (ETL) e visualização de dados em uma ferramenta OLAP. Durante o desenvolvimento foi utilizado o *Microsoft Excel* para análise prévia dos dados, o *SQL Power Architect* para criação do modelo dimensional, o banco de dados Microsoft SQL Server para o armazenamento de dados do *Staging* e do DW, o SQL Server Management Studio para o gerenciamento, administração e consulta dos componentes do Microsoft SQL Server, o Pentaho Data Integration como ferramenta de ETL e o Microsoft Power BI como ferramenta de visualização de dados. Os resultados obtidos no trabalho desenvolvido foram válidos, pois foi possível criar análises relevantes utilizando conceitos de BI.

Palavras-chave: *Business Intelligence*, *Data Warehouse*, ETL, Judiciário, Dados Abertos.

LISTA DE SIGLAS

BI - Business Intelligence

CNJ – Conselho Nacional de Justiça

DM - Data Mart

DPJ - Departamento de Pesquisas Judiciárias

DAX - Data Analysis Expression

DW - Data Warehouse

ETL – Extract, Transform and Load

IBGE – Instituto Brasileiro de Geografia e Estatística

OLAP – Online Analytical Processing

PDI - Pentaho Data Integration

SIESPJ - Sistema de Estatísticas do Poder Judiciário

SQL – Structured Query Language

SSMS – SQL Server Management Studio

TJ - Tribunal de Justiça

VBA – Visual Basic Application

LISTA DE QUADROS

Quadro 1 – Tabela dimensão_produto.....	24
Quadro 2 – Tabela fato_venda.....	24
Quadro 3 - Regras de Negócio	45
Quadro 4 - Requisitos Funcionais	47
Quadro 5 - Requisitos não Funcionais.....	48
Quadro 6 – Processo de Desenvolvimento	57
Quadro 7 – Arquivos Fonte	58
Quadro 8 – Campos Power BI.....	72
Quadro 9 – Perguntas Estratégicas x Elementos Gráficos	76

LISTA DE FIGURAS

Figura 1 – Dado, Informação, Conhecimento	18
Figura 2 – Data Mart Dependente	21
Figura 3 – Data Mart Independente.....	21
Figura 4 – Esquema Estrela	25
Figura 5 – Esquema Flocos de Neve	27
Figura 6 – Processo ETL	29
Figura 7 – Drill Down	32
Figura 8 – Drill Up	33
Figura 9 – Slice.....	34
Figura 10 – Etapas Metodológicas	41
Figura 11 – Modelagem Dimensional	49
Figura 12 – Quadrante Mágico da Gartner – Analytics and Business Intelligence Plataforma.....	53
Figura 13 – Processo Desenvolvimento	57
Figura 14 – Transformação Staging Indicadores.....	60
Figura 15 – Transformação Staging Documentacao	61
Figura 16 – Transformação dPorte	61
Figura 17 – Select Input dPorte	62
Figura 18 – Transformação dTempo	62
Figura 19 – Select Input dTempo	63
Figura 20 – Select Input dTribunal.....	63
Figura 21 – Select Input dTribunal 2.....	63
Figura 22 – Transformação dTribunal.....	64
Figura 23 – Select Input dIndicador	65
Figura 24 – Select Input dIndicador 2	65
Figura 25 – Transformação dIndicador	65
Figura 26 – Select Input fIndicador	66
Figura 27 – Transformação fIndicador.....	67
Figura 28 – Job Staging	68
Figura 29 – Job DW.....	68
Figura 30 – Tela Inicial do Sistema Desktop	70
Figura 31 – Tela Inicial do Sistema Mobile	71
Figura 32 – Filtro Indicador.....	73
Figura 33 – Cards com Indicador Filtrado.....	74
Figura 34 – Cards com Indicador e Tribunal Filtrados	74
Figura 35 – Gráficos	75
Figura 36 – Operações OLAP	76

SUMÁRIO

1	INTRODUÇÃO	11
1.1	PROBLEMÁTICA	12
1.2	OBJETIVOS	13
1.2.1	Objetivo Geral	14
1.2.2	Objetivos Específicos	14
1.3	JUSTIFICATIVA	14
1.4	ESTRUTURA DA MONOGRAFIA	15
2	BUSINESS INTELLIGENCE	16
2.1	DADOS, INFORMAÇÕES E CONHECIMENTO	17
2.2	DATA WAREHOUSE	18
2.2.1	Data Mart	20
2.2.2	Tabela Fato	22
2.2.3	Tabela Dimensão	23
2.2.4	Esquema Estrela (Star Schema)	25
2.2.5	Esquema Flocos de Neve (Snowflake Schema)	26
2.2.6	Granularidade	27
2.2.7	Staging Area	28
2.3	ETL	29
2.4	OLAP	30
2.4.1	Operações	31
2.4.1.1	Drill Down e Drill Up	32
2.4.1.2	Drill Across	33
2.4.1.3	Drill Throught	34
2.4.1.4	Slice and Dice	34
2.4.2	Medidas	35
2.5	JUDICIÁRIO	35
2.5.1	Tribunais	36
2.5.2	CNJ	36
2.5.2.1	Relatório Justiça em Números	37
3	METODOLOGIA	39
3.1	CARACTERIZAÇÃO DO TIPO DE PESQUISA	39
3.2	ETAPAS METODOLÓGICAS	41
3.3	DELIMITAÇÕES	42
4	MODELAGEM	43
4.1	REQUISITOS	43
4.1.1	Perguntas estratégicas	43
4.1.2	Regras de negócio	45
4.1.3	Requisitos Funcionais	46
4.1.4	Requisitos não Funcionais	47
4.2	MODELAGEM DIMENSIONAL	48
4.2.1	Dimensão Tempo	49
4.2.1	Dimensão Tribunal	50
4.2.2	Dimensão Indicador	50
4.2.3	Dimensão Porte	50
4.2.4	Fato Indicadores	51
5	DESENVOLVIMENTO	52

5.1	FERRAMENTAS UTILIZADAS	52
5.1.1	Microsoft Power BI.....	52
5.1.2	Microsoft SQL Server.....	54
5.1.3	SQL Server Management Studio	54
5.1.4	Pentaho Data Integration Kettle	54
5.1.5	Microsoft Excel.....	55
5.1.6	SQL Power Architect.....	56
5.1.7	DAX.....	56
5.2	EXPERIMENTO	56
5.2.1	Arquivos.....	58
5.2.2	Banco de Dados	59
5.2.3	Transformação Staging	60
5.2.4	Transformação dPorte.....	61
5.2.5	Transformação dTempo	62
5.2.6	Transformação dTribunal.....	63
5.2.7	Transformação dIndicador	64
5.2.8	Transformação fIndicador	66
5.2.9	Job Staging	67
5.2.10	Job DW	68
5.2.11	Protótipo	69
5.2.11.1	Filtros	72
5.2.11.2	Cards	73
5.2.11.3	Gráficos Principais	75
5.3	AVALIAÇÃO ENTREVISTA STAKEHOLDER.....	77
5.4	CONSIDERAÇÕES	77
6	CONCLUSÕES E TRABALHOS FUTUROS	79
6.1	CONCLUSÃO.....	79
6.2	TRABALHOS FUTUROS.....	80
	REFERÊNCIAS	82

1 INTRODUÇÃO

O acesso aos dados governamentais está cada vez mais disponível para a população em geral, porém mesmo com essa facilidade, não são dados simples de serem interpretados por diversos motivos, como por exemplo a estrutura em que se encontram ou nomenclaturas de difícil entendimento.

Em janeiro de 2009, em um de seus primeiros atos, o então presidente dos Estados Unidos, Barack Obama, emitiu um memorando presidencial sobre “*Transparency and Open Government*”, no qual estabelece a transparência, participação e colaboração entre todos. Foi considerado uma das pedras angulares para um governo aberto e foi referência para muitos outros países. (HARRISON et al., 2012).

Conforme *Open Government Partnership* (2011) os dados abertos governamentais são essenciais para:

- a) aumentar a disponibilidade de informações sobre atividades governamentais;
- b) apoio na participação cívica;
- c) implementar os mais altos padrões de integridade profissional em todas as administrações;
- d) aumentar o acesso às novas tecnologias para a transparência e prestação de contas.

Seguindo essa tendência, o Brasil também iniciou esforços para disponibilizar os dados governamentais, de acordo com o Governo Digital (2016):

Os Dados Abertos Governamentais são uma metodologia para a publicação de dados do governo em formatos reutilizáveis, visando o aumento da transparência e maior participação política por parte do cidadão, além de gerar diversas aplicações desenvolvidas colaborativamente pela sociedade. [...] No caso do Brasil, vários órgãos da Administração Pública já publicam seus dados na web, na forma de relatórios e balanços para que os cidadãos possam acompanhar os resultados das ações de governo.

Segundo a Lei Complementar 131 sancionada pelo ex-presidente da república, Luiz Inácio Lula da Silva em 27 de maio de 2009 (BRASIL, 2009), conhecida como Lei da Transparência, assegura ao cidadão que o estado, município e a União divulgarão dados na Internet em tempo real “[...] liberação ao pleno conhecimento e acompanhamento da sociedade, em tempo real; de informações pormenorizadas sobre a execução orçamentária e financeira, em meios eletrônicos de acesso público”.

Complementando a Lei da Transparência, existe também a Lei de Acesso à Informação (LAI) que:

[...] destinam-se a assegurar o direito fundamental de acesso à informação e devem ser executados em conformidade com os princípios básicos da administração pública e com as seguintes diretrizes:
I - observância da publicidade como preceito geral e do sigilo como exceção;
II - divulgação de informações de interesse público, independentemente de solicitações;
III - utilização de meios de comunicação viabilizados pela tecnologia da informação;
IV - fomento ao desenvolvimento da cultura de transparência na administração pública;
V - desenvolvimento do controle social da administração pública.

Conforme Klein, Luciano e Macadar (2015), esta é uma lei que permite que as pessoas tenham acesso aos dados do Estado; existem dois métodos de acesso, ativo e passivo. Na opção ativa, o governo oferece proativamente o acesso às informações de interesse geral. No caso da opção passiva, a pessoa física ou jurídica deve solicitar informações específicas.

É possível notar que essas leis asseguram aos cidadãos o acesso e o uso da política dos dados abertos governamentais, com isso, potencialmente melhorar a eficiência, efetividade e inovação na área governamental como um todo. Além disso, possui a vantagem de diminuir as perguntas frequentes para figuras públicas e serem respondidas de forma mais rápida e assertiva via análise dos dados (UBALDI, 2013).

Neste trabalho são utilizados os dados abertos fornecidos pelo portal do Conselho Nacional de Justiça (CNJ). No arquivo os dados não estão padronizados e possui colunas com nomes que não permitem um entendimento rápido pelos analistas de negócios. Na próxima seção e capítulos serão apresentadas as explicações e razões do motivo dessa estrutura não ser a ideal para uma análise de dados mais aprofundada.

1.1 PROBLEMÁTICA

Conforme o Conselho Nacional de Justiça (CONSELHO NACIONAL DE JUSTIÇA, 2017) o Relatório Justiça em Números é a principal fonte das estatísticas oficiais do Poder Judiciário e mostra com detalhamentos a estrutura e a litigiosidade, além dos indicadores e análises para a gestão judiciária brasileira.

O CNJ disponibiliza os dados para o público, ou seja, são dados que todos podem utilizar, porém estão desnormalizados. Para Sanders e Shin (2001, p. 1, tradução nossa) "a desnormalização tem sido utilizada em muitas implementações estratégicas de banco de dados para aumentar o desempenho do banco de dados e reduzir os tempos de resposta [...]", ou seja, essa solução é boa para performance, porém para análise de dados é inviável:

O objetivo da normalização é organizar os dados em estruturas estáveis e assim minimizar as anomalias na hora da atualização e maximizar a acessibilidade de dados. Embora a normalização seja geralmente considerada como regra para o modelo relacional, ainda existe momentos em que arquitetos de banco de dados utilizam a desnormalização para aumentar a performance [...] normalização resulta em muitos benefícios. (SANDERS; SHIN, 2001, p.1, tradução nossa)

Além disso, como dito na introdução, os nomes das colunas dos indicadores disponibilizados no arquivo estão em um padrão que não significa nada para um analista de negócio ou partes interessadas, como por exemplo: "A1", "A2", "A3", dificultando a compressão.

Para o público geral, os dados por si só não significam muita coisa. É preciso que estejam organizados em uma estrutura legível, clara e de fácil compreensão. Conforme Menninger, Westland e Date (1995, 1992, 1997 apud ZAKER et al, 2009), os dados devem ser organizados em uma estrutura bem equilibrada através da normalização com o fim de otimizar a acessibilidade dos dados, mas esse procedimento implica certas deficiências que resultam em um menor desempenho do sistema.

1.2 OBJETIVOS

Nesta seção são descritos o objetivo geral e os objetivos específicos deste trabalho.

1.2.1 Objetivo Geral

Desenvolver uma solução de *Business Intelligence* com indicadores gerados a partir de dados abertos da Justiça.

1.2.2 Objetivos Específicos

- Modelar uma base de dados dimensional para ser utilizada como base da solução de BI.
- Desenvolver uma solução para transformação e carga da modelagem dimensional.
- Aplicar uma solução de *Business Intelligence* para visualizar os indicadores gerados a partir da base dimensional.

1.3 JUSTIFICATIVA

Visto a importância dos dados governamentais, a ministra Cármen Lúcia reforça a criticidade de dados dos Tribunais de Justiça (SUPREMO TRIBUNAL FEDERAL, 2017) confiáveis dizendo que não é possível oferecer transparência para os cidadãos se não existir informações corretas. A ministra ainda relata que no ano de 2017, 7 dos 27 Tribunais de Justiça não forneceram dados para o Conselho Nacional de Justiça, dificultando assim, ter um panorama exato da situação do Judiciário.

De acordo com o Marcelo Guedes (CONSELHO DA JUSTIÇA FEDERAL, 2015), presidente da Associação Brasileira de Jurimetria, uma boa política pública deve levar em consideração os dados que realmente representam a realidade, após isso, serem analisados e assim, poder unir esforços para a reflexão dos números. Além disso, comenta que quanto mais a população estiver desenvolvida, educada e munida de informações, mais poderá recorrer aos

tribunais para a solução de problemas, com isso, ajudando na construção de uma sociedade mais justa e democrática.

Dito isso, é entendido que dados públicos, estatística e análise de dados é fundamental para a verdadeira democratização do país. Porém, por mais que estejam disponíveis, esses dados são difíceis de serem interpretados e entendidos pela população em geral. Vendo esse contexto este trabalho tem como o objetivo facilitar o acesso à informação dos dados judiciais fornecidos pelo Conselho Nacional de Justiça (CNJ), trazendo mais informações para a população e consequentemente fomentando ainda mais a cidadania e com isso tornando-se uma parte ativa à democracia.

1.4 ESTRUTURA DA MONOGRAFIA

No capítulo 1 é apresentado a Introdução, Problemática, Objetivo Geral, Objetivos Específicos e Justificativa.

No capítulo 2 é apresentado a revisão bibliográfica, com teorias de modelagem de dados, modelagem relacional, modelagem dimensional, modelo estrela, visualização de dados, teoria de “*extract, transform and load*” (ETL), teorias específicas de “*Business Intelligence*” (BI).

No capítulo 3 é mostrado o planejamento do trabalho, com as delimitações, caracterização do tipo de pesquisa, etapas metodológicas e proposta de solução.

No capítulo 4 é descrito a modelagem das tabelas, seguindo o modelo estrela para o desenvolvimento.

No capítulo 5 é apresentado o desenvolvimento da solução com *Business Intelligence*, análise de dados e resultado do questionário realizado com especialistas na área do negócio.

No capítulo 6 é apresentado conclusões, resultados obtidos e trabalhos futuros.

2 BUSINESS INTELLIGENCE

Nas palavras de Luhn (1958), um dos primeiros a discorrer sobre *Business Intelligence* (BI), define *business* como uma coleção de atividades realizadas para qualquer fim, seja ciência, tecnologia, comércio, indústria e entre outros. Já *intelligence*, em um modo geral, como a capacidade de aprender as interrelações dos fatos apresentados para guiar ações em busca de um objetivo desejado.

Business Intelligence é definido pela utilização de variadas fontes de informação para definir estratégias de competitividade nos negócios da empresa. Podem ser incluídos nessa definição os conceitos de estruturas de dados, representadas pelos bancos de dados tradicionais, *data warehouse* e *data mart*, criados objetivando o tratamento relacional e dimensional de informações (BARBIERI, 2011, p.95).

Para Primak (2008, p.1) o termo BI, não é recente como se imagina:

O seu conceito prático já era usado pelo povo antigo. A sociedade do Oriente Médio antigo utilizavam os princípios básicos do *Business Intelligence* quando cruzavam informações obtidas junto à natureza em benefício de suas aldeias. Analisar comportamento das marés, os períodos chuvosos e de seca, a posição dos astros, entre outras, eram formas de obter informações que eram usadas para tomar decisões importantes que permitissem a melhoria de vida de suas respectivas comunidades.

Em 1989, Howard Dresner, que mais tarde passa a ser analista do Gartner Group, propôs BI como um termo abrangente para descrever conceitos e métodos para melhorar a tomada de decisões empresariais utilizando sistemas baseados em acontecimentos (POWER, 2007).

Turban e outros (2009, p.21) ressaltam a importância de empresa utilizarem dados com qualidade para obter-se vantagem competitiva:

As empresas, privadas ou públicas, sentem crescentes pressões forçando-as a responder rapidamente a condições que estão sempre mudando, além de terem que ser inovadoras na maneira com que operam. Essas atividades exigem das empresas agilidade, tomadas de decisão rápidas e frequentes, sejam elas estratégicas, táticas e operacionais, algumas das quais são muito complexas. Tomar essas decisões pode exigir quantidades consideráveis de dados oportunos e relevantes, além de informações e conhecimento. O processamento dessas informações, na estrutura das decisões necessárias, deve ser feito de forma rápida, com frequência em tempo real, e comumente exige algum apoio computadorizado.

No texto de Turban e outros (2009, p.22) também traz um interessante case da Toyota Motor Sales USA:

A empresa compra os carros na fábrica da Toyota no Japão, toma posse dos veículos e então os vende para os revendedores da Toyota nos Estados Unidos. [...] o custo para a empresa era de US\$ 144 a 160 milhões por ano pelo transporte. A gerência da Serviços Logísticos da Toyota exige rastreamento preciso para assegurar que os carros certos vão para os revendedores certos a tempo. [...] a solução veio então com um

sistema que permite que os executivos vejam áreas que merecem atenção em suas unidades de negócio e investiguem mais para identificar os problemas com exatidão, bem como as suas causas. [...] Dentro de poucos dias, o sistema começou a apresentar resultados fascinantes. Por exemplo, o sistema ajudou a descobrir que a Toyota era cobrada duas vezes por um envio especial por trem (um erro de US\$ 800.000).

Na compreensão de Primak (2008, p.1) o atual interesse pelo BI vem crescendo assustadoramente na medida em que seu emprego possibilita às organizações realizar uma série de análises e projeções, de forma a agilizar os processos relacionados às tomadas de decisão.

Piton (2018, p.18) ainda afirma que “ninguém mais toma decisão porque acordou feliz ou triste. Hoje, em qualquer empresa, as decisões são tomadas com base em dados e evidências. Nós produzimos uma massa de dados gigante, e esses dados são uma mina de ouro, eles precisam ser lapidados”.

Importância do BI é ressaltada por Piton (2018) como um processo muito mais complexo do que a maioria das pessoas pensam. Possui o poder de fechar uma empresa, demitir, aumentar faturamento e garantir uma tomada de decisão mais precisa.

2.1 DADOS, INFORMAÇÕES E CONHECIMENTO

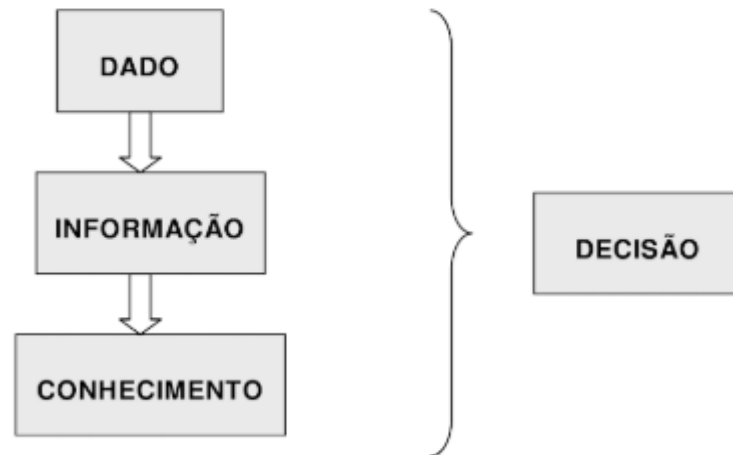
Angeloni (2003) afirma que há dificuldade em se diferenciar dado, informação e conhecimento. Mesmo sendo os principais elementos na comunicação e tomada de decisão, beneficiar-se destes, não é um processo tão simples como parece. Angeloni (2003, p.18) ressalta que: “existem informações que os indivíduos não percebem e não vêem; informações que vêem, e não ligam; informações que vêem, e não entendem ou não decodificam; informações que vêem e usam; informações que procuram; informações que adivinham”.

Muitas empresas estão utilizando não apenas os seus dados operacionais e transacionais, mas também os disponíveis na web, como por exemplo, textos publicados em microblogs, como o Twitter, para identificar a opinião de um grupo sobre um dos seus produtos ou serviços. Esses dados disponíveis na web combinados com os internos da organização geram informações ainda mais relevantes e estratégicas para a etapa da tomada de decisão (CECI, 2012, p.16).

Dado é um modelo de algo que, sozinho, não traz nenhum sentido ou gera sentido duplo. O valor 24/05/1975 não quer dizer muita coisa, no entanto, ao dizer que se trata de uma data de nascimento, gera-se o conhecimento de uma data de aniversário e consequentemente, a possibilidade de tomar uma decisão, como dar um presente (PRIMAK, 2008).

Um dado é a estrutura fundamental sobre a qual um sistema de informação atua. A informação pode ser vista como uma representação ordenada e enxuta dos dados resultantes de uma consulta que permite a visualização e interpretação deles. O conhecimento provém da interpretação das informações apresentadas pelo sistema de banco de dados (CECI, 2012, p.98).

Figura 1 – Dado, Informação, Conhecimento



Fonte: Primak (2008, p. 12).

Como é possível ver na figura 1, Primak (2008) reforça que informação é a base para construção do conhecimento. Portanto, a informação não é conhecimento, mas sim parte do processo deste.

2.2 DATA WAREHOUSE

Conforme Inmon e Kimball (2009, 1998), entender o significado, a arquitetura, seu papel e propósito é essencial para a compreensão e trabalho com BI justamente por ser a espinha dorsal e o coração da estrutura de apoio a decisão.

De acordo com Inmon (2009), um dos objetivos principais do *data warehouse* (DW) é deixar o trabalho dos analistas que irão utilizar os dados mais simples do que utilizar um ambiente legado clássico, pelo motivo de existir somente uma fonte de dados, integrando tudo em um lugar único, formando uma base clara para facilitar a reutilização e conciliação dos dados.

Um *data warehouse* é uma coleta de dados orientada por assuntos, integrada, não volátil e variante no tempo, para apoio às decisões da administração. O *data warehouse* contém dados corporativos granulares. Os dados do *data warehouse* podem ser usados para muitos propósitos diferentes, incluindo dados não utilizados,

esperando por requisições futuras que são desconhecidos hoje (INMON, 2009, p.31, tradução nossa).

Essa estrutura deve fornecer as visões mais importantes e relevantes dos dados corporativos, com isso é preciso prover possibilidade e flexibilidade de como os dados serão interpretados e analisados posteriormente, produzindo uma fonte estável de informações históricas, de uma forma consistente e confiável para qualquer consumidor autorizado (IMHOFF et al, 2003)

Desde o começo da estruturação, é preciso ter em mente que existe uma enorme necessidade de se guardar informações históricas e que por esse motivo o DW pode crescer de uma forma considerável, então a arquitetura deve ser construída de uma forma escalável, para poder acomodar o aumento das informações de uma forma eficiente. Além disso, é preciso ser feito com o intuito de viabilizar para qualquer forma de tecnologia ou ferramenta utilizada na empresa, assim possibilitando até mesmo criar outros *data marts* paralelos a partir dos dados contidos no *data warehouse* (IMHOFF et al, 2003)

A característica mais importante de um *data warehouse* de acordo com o Inmon (2005) é que ele deve ser integrado, como já dito, os dados são alimentados de várias fontes de dados, como por exemplo: .csv, .xlsx, JSON, banco de dados relacionais e outros. Esses dados são padronizados, renomeados, convertidos, resumidos, numeradas, esse processo é conhecido como *Extract, Transform and Load*, comumente chamando de ETL, esse tópico será discutido nos próximos capítulos.

Kimball (1998) faz uma metáfora interessante sobre a arquitetura do DW, dizendo que é inimaginável pensar em construir uma casa sem um conjunto de plantas, porém é comum pessoas começarem um projeto de DW, que pode custar muito mais caro, sem ter ideia clara do que querem construir, é preciso ter um planejamento prévio para entender o real trabalho que irá ser feito.

Ainda de acordo com o Kimball (1998), uma arquitetura de boa qualidade agrega valor da mesma forma dos projetos de construção civil, aumentando significativamente a eficácia e a flexibilidade do sistema como um todo, os valores principais segundo ele são:

- a) Comunicação: fornece uma maneira de ajudar os gerentes a entender a complexidade do projeto. A própria arquitetura se transforma em uma ferramenta de comunicação, ajudando os participantes do processo a criar um senso de onde eles se encaixam e o que precisam realizar.

- b) Planejamento: muitos detalhes arquitetônicos e técnicos acabam se perdendo em um plano de projeto, a arquitetura traz tudo à tona e descreve como se encaixam.
- c) Flexibilidade e manutenção: após planejar, é possível prever o maior número possível de problemas e construir um sistema que possa lidar com isso, criando um processo para navegar e manter eficientemente.
- d) Aprendizagem: para acontecer a aprendizagem é necessária uma boa documentação do sistema, como já dito anteriormente, ela é útil para diversos casos, por exemplo, auxiliar os novos membros da equipe à adaptação. Criar sozinho um mapa mental da arquitetura via tentativa e erro é perigoso e passível de muitos erros, além de não ser tão produtivas quanto seriam se existisse uma documentação precisa.
- e) Produtividade e reuso: as ferramentas e metadados são fortes aliadas que possibilitam a melhoria na produtividade e reutilização, pois com o conhecimento da arquitetura e dos dados fica mais fácil reutilizar processos existentes em vez de construir algo complexo do zero.

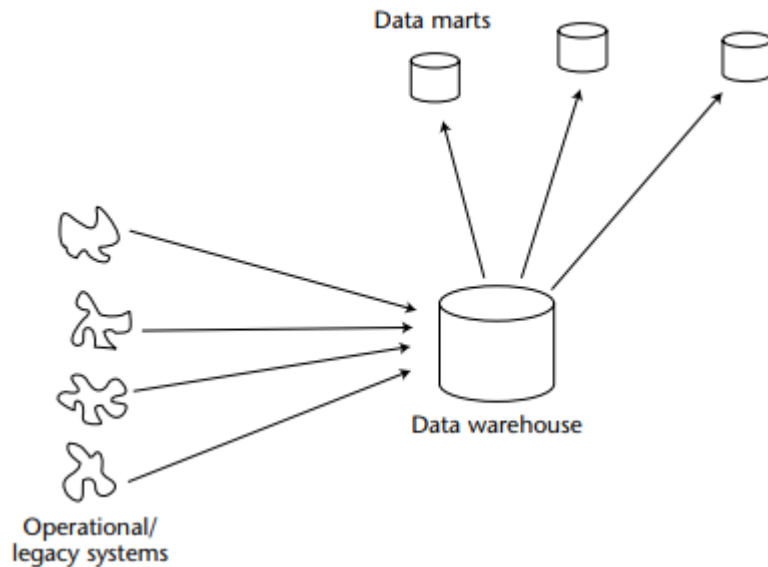
Diversas vezes, o BI é confundido com o DW. Para esclarecer, o último está incorporado em um projeto de *Business Intelligence*. Sintetizando, os objetivos principais de um DW é ter dados históricos pronto para análise, ter um bom tempo de resposta nos *selects*, garantir consistência e confiabilidade dos dados, ser simples e fácil a interpretação dos dados, padronizar a nomenclatura e dados de acordo com as regras pré-definidas, ter um repositório único e centralizado. Se esses tópicos forem garantidos, significa que a técnica e a modelagem foram bem elaboradas (PITON, 2018).

2.2.1 Data Mart

Kimball (1998) explica que o *data mart* é uma fatia do *data warehouse*, ou seja, um DW é basicamente formado pela união de todos os seus *data marts*. Trabalhar desta forma é interessante, pois dividimos um problema complexo e grande que levaria meses ou anos para

ser criado em vários pequenos pedaços, além disso, é possível criar marcos de entregas, patrocinadores e responsáveis diferentes para cada área, ajudando a diminuir o custo do projeto.

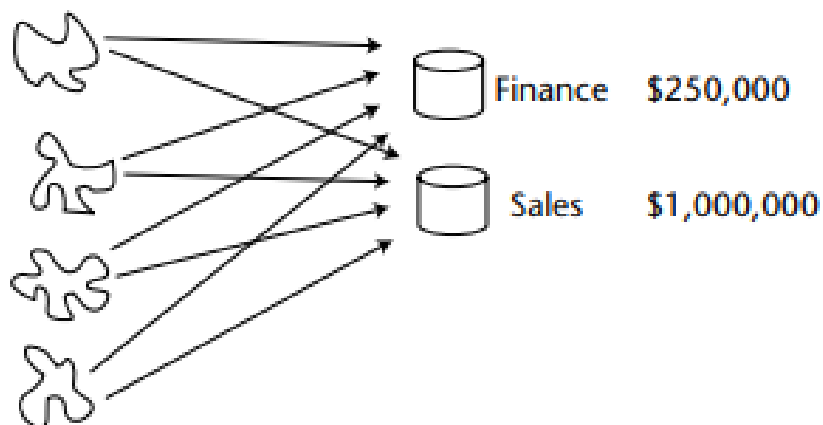
Figura 2 – Data Mart Dependente



Fonte: Inmon (2009, p. 205).

Como mostrado nas figuras 2 e 3, em relação aos *data marts*, sob o ponto de vista do Inmon (2009), podem existir os dependentes e os independentes. Um *data mart* dependente é criado a partir de um DW, ou seja, a fonte de dados é o DW, para esse modelo é necessário que a equipe pense globalmente, além do mais, comumente a arquitetura costuma ficar muito complexa para a manutenção e requer um planejamento antecipado.

Figura 3 – Data Mart Independente.



Fonte: Inmon (2009, p. 205).

O *data mart* independente, mostrado na figura anterior, pode ser criado por demanda, se por algum motivo ocorrer problema em um deles, não irá afetar o todo, caindo

bastante a complexidade e dessa forma, garantindo a estabilidade e a consistência (INMON, 2009).

2.2.2 Tabela Fato

Um dos pontos mais importantes da modelagem dimensional é que praticamente todos tipos de dados de negócios podem ser representados por um cubo de dados, onde as bordas são definidas as dimensões e as células do cubo são os valores medidos (KIMBALL, 1998).

Conforme Kimball (2008, tradução dos autores)

Uma tabela de fatos contém as medidas numéricas produzidas por um evento de medição operacional no mundo real. No menor grão, uma linha da tabela de fatos corresponde a um evento de medição e vice-versa. Assim, o projeto fundamental de uma tabela de fatos é inteiramente baseado em uma atividade física e não é influenciado pelos relatórios eventuais que podem ser produzidos. Além das medidas numéricas, uma tabela de fatos sempre contém chaves estrangeiras para cada uma de suas dimensões associadas, bem como chaves de dimensão degeneradas opcionais e selos de data / hora. As tabelas de fatos são o principal alvo de cálculos e agregações dinâmicas decorrentes de consultas.

Em outras palavras, “a tabela fato armazena o que ocorreu, é o fato propriamente dito, por isso ela tem esse nome” (PITON, 2018, p, 101). Normalmente, os fatos na área de negócios são numéricos e por isso, o arquiteto do DW deve suspeitar que se houver valores numéricos, é provavelmente um fato (KIMBALL, 1998).

Ainda conforme Kimball (2008) as tabelas fatos podem ser agrupadas em três tipos: granularidade transacional, granularidade *snapshot* e granularidade *snapshot* cumulativa, cada um com suas respectivas características.

- a) Granularidade transacional: é medida em um único instante, as medidas, valores e a informação são só válidos para aquele instante e contexto específico, normalmente é uma das tabelas fatos que possui o maior número de registros.
- b) Granularidade *snapshot*: corresponde a um intervalo de tempo pré-definido, diferente da transacional, as informações resumem as informações durante ou ao final do intervalo de tempo. É uma das tabelas mais utilizadas para se

criar um relatório. Um dos pontos positivos é ser possível de garantir que todas as chaves estrangeiras estejam presentes.

- c) Granularidade *snapshot* cumulativa: ela é usada quando um evento possui um início e fim bem definidos, por exemplo, em matrículas de alunos, registros de *call centers*, processamento de pedidos. Essa tabela é revisada e reescrita à medida que o evento progride através das suas etapas de início, meio e fim. Normalmente o *snapshot* cumulativo é muito menor do que as outras tabelas, por utilizar a estratégia de reescrita.

Nas concepções de Ceci e Kimball (2012, 1998) é importante saber que para cada situação a tabela será diferente, ela deve responder as perguntas de negócio dos usuários chaves, o objetivo do DW deve estar muito claro antes de qualquer coisa, identificar os assuntos, granularidade e a complexidade é essencial para um funcionamento eficiente, pois por exemplo, quanto menor for o grão dos dados em uma tabela fato, maior será a quantidade de dados e detalhes.

2.2.3 Tabela Dimensão

Após entender o que é um fato e a sua granularidade é preciso saber com o qual informação ela será agrupada e cruzada. A tabela dimensão possui elementos que participam da tabela fato, determinando um contexto do assunto, contendo as características do evento (KIMBALL, 1998; INMON, 2005; CECI, 2012).

Nas palavras de Ceci (2012, p.89) uma boa forma de começar a criar as dimensões é: “após identificar os ‘assuntos’ que o DW irá tratar, devem-se levantar as dimensões que são utilizadas para agrupar e cruzar as informações contidas no DW, para isso, devemos responder a questões do tipo: ‘quem?’, ‘quando?’, ‘o quê?’ e ‘onde?’”.

Por exemplo, se a fato for a venda de um produto, pode-se perguntar ‘quem’ realizou a venda, ‘quando’ ocorreu, ‘o que’ foi vendido e ‘onde’ a venda ocorreu, com isso, é possível ver que nesse exemplo simples terá quatro dimensões: de vendedor, tempo, produto e local.

Quadro 1 – Tabela dimensão_produto

Produto	Desc	Categoria
S1	Cerveja Sk	Alcoólica
S2	Coca-Cola	Refri
S3	Pepsi-Cola	Refri
S4	Cerveja black	Alcoólica
S5	Espumante	Alcoólica

Fonte: Barbieri (2011, p. 276).

Como explicado por Kimball Group (2008) e Kimball (1998) e exemplificado no quadro 1 e 2, diferente da tabela fato onde a linha se torna única baseado em um conjunto de chaves estrangeiras da dimensão, os dados da tabela dimensão deverão ser únicos, como mostrado na imagem anterior.

Quadro 2 – Tabela fato_venda

Produto	Loja	Data	Qtd	Custo-Unitário
S1	S2	S1	127	10,00
S1	S2	S2	65	20,00
S2	S2	S1	32	2,00
S3	S2	S2	250	1,00
S3	S3	S2	26	3,00

Fonte: Barbieri (2011, p. 276).

Resumindo, sob o ponto de vista de Kimball, Piton, Barbieri e Machado (1998, 2018, 2011, 2013), a dimensão possui os atributos descritivos usados pelo BI, filtrando e agrupando os fatos. As dimensões podem ser consideradas a alma do DW, pois só com ela é possível dar sentido aos dados, criando análises importantes para o negócio, sendo impulsionadores da experiência do BI como um todo para os seus usuários, independente de quem sejam.

2.2.4 Esquema Estrela (Star Schema)

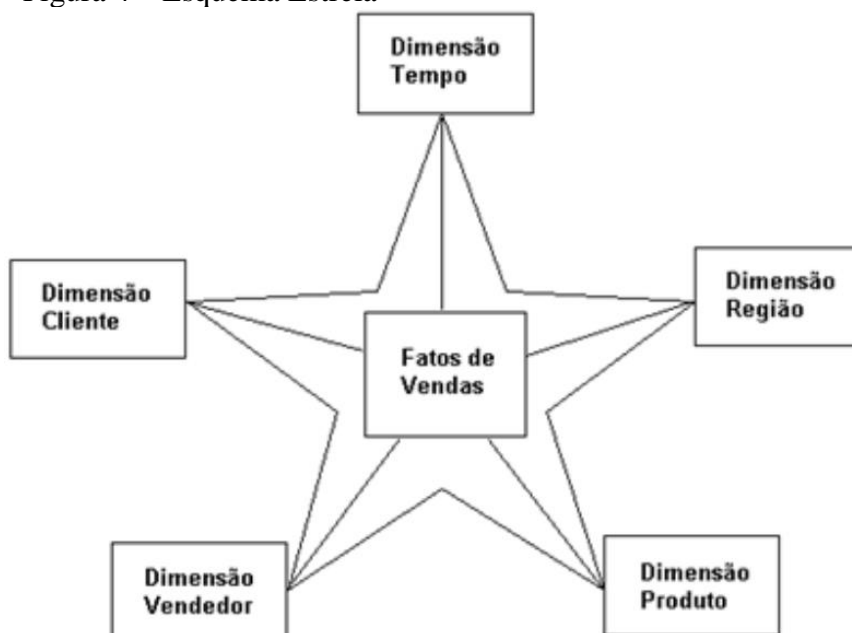
As tabelas fato são cercadas por um conjunto de tabelas de dimensões descrevendo, precisamente, o que é conhecido de cada linha de registro da tabela. Por causa da estrutura característica de um modelo dimensional, é frequentemente chamado de Esquema Estrela (KIMBALL, 2002). Abaixo na compreensão de Barbieri:

A abordagem de Ralph Kimball, aponta para projetos de *data marts* separados, que deverão ser integrados na medida da sua evolução. Os projetos serão menores, independentes, focando áreas ou assuntos específicos e terão a sua conexão com o passar do tempo, desde que mantida a compatibilidade dimensional entre as chaves das tabelas. (BARBIERI, 2013, p.113)

Pujari (2001) explica que o esquema estrela é um paradigma de modelagem em que o DW contém uma grande tabela fato central única e um conjunto de tabelas dimensões menores, uma para cada dimensão, cada uma única e altamente desnormalizada, conforme a figura 4. Possuindo as seguintes vantagens:

- a) Fácil de entender.
- b) Fácil de definir hierarquias.
- c) Reduz o número de *joins* físicos.
- d) Requer baixa manutenção.
- e) Metadados muito simplificados.

Figura 4 – Esquema Estrela



Fonte: Machado (2013, p. 92).

Conforme Barbieri (2012) aponta como desvantagem a possibilidade de produzir diversos *data mart* sem uma perfeita ligação entre eles, além de uma provável duplicação de esforços na fase de extração, preparação e carga de dados, fase mais conhecida como ETL, que será explicado nos próximos capítulos.

Para Piton (2018) o Esquema Estrela é mais fácil de implementar e acabou sendo mais utilizado porque além de hoje em dia o espaço em disco ser muito barato, ele permite entregar projetos em pequenas partes. É possível criar um assunto ou departamento em uma estrela para só depois projetar outro, e assim por diante.

2.2.5 Esquema Flocos de Neve (Snowflake Schema)

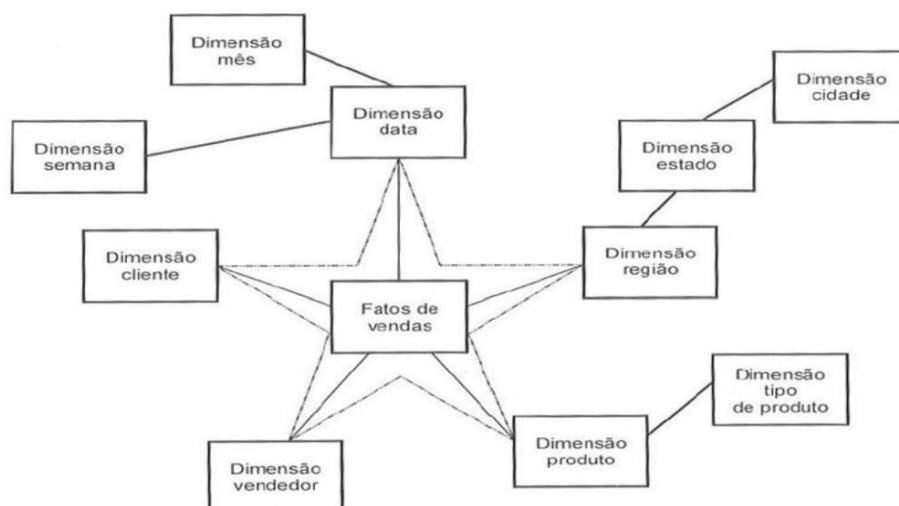
Para Barbieri (2012, p.113) a ênfase do esquema flocos de neve, proposto por Inmon, sempre foi um “grande depósito central de informações empresariais tratadas, limpas e integradas, construído inicialmente, e de onde outros depósitos secundários são originados e construídos”.

O texto de Adamson (2006, p.284, tradução nossa) compara o esquema estrela com o esquema flocos de neve no que diz respeito a dimensões e fatos, conforme é possível ver na figura 5: “o que o difere do esquema estrela é que no esquema flocos de neve, as tabelas dimensão são normalizadas, tornando explícitas as hierarquias entre os atributos dimensionais”.

Barbieri (2012, p.113) realça o enfoque da abordagem de Inmon: “se concentrou inicialmente no estilo mais tradicional de construção de bancos de dados, muito próximo daquele sugerido nos primeiros projetos de bancos de dados, nos quais se buscava uma forte integração entre todos os dados da empresa que habitavam áreas funcionais diferentes”.

Uma vantagem do modelo flocos de neve é que as tabelas de dimensão são mais versáteis. Por exemplo, se uma dimensão de tempo é normalizada em várias tabelas de dimensão, uma tabela de ano, uma tabela de mês e uma tabela de data, em seguida, é possível ter várias tabelas de fatos que armazenam dados a outro nível de granularidade com respeito ao tempo; uma tabela de fatos pode estar relacionada com a tabela de data, enquanto outro podia se relacionar a tabela de mês ou o ano. Como outro exemplo, se o depósito contém uma dimensão de cliente e uma dimensão Geografia, então a dimensão Geografia pode ser usada para descrever fatos diretamente, por exemplo, o fato ocorreu em uma geografia específica, e também pode ser usado para descrever os clientes, por exemplo, um cliente reside em um local específico (VELTEM, 2012).

Figura 5 – Esquema Flocos de Neve



Fonte: Machado (2013, p. 94).

Embora reduza o tamanho de dimensões, Adamson (2006) afirma que o esquema flocos de neve não economiza quantidade de espaço em disco, isso ocorre porque as maiores tabelas são, geralmente, as tabelas fatos. Além disso, as consultas incluirão mais *joins*, que são necessárias para relacionar as numerosas tabelas dimensões, podendo ter um impacto no desempenho.

2.2.6 Granularidade

Para Inmon e Machado (2005), a parte mais importante da modelagem de um DW é a definição da granularidade. A questão da granularidade circula toda a arquitetura que envolve o ambiente do DW. Granularidade refere-se ao nível de detalhe ou resumo das unidades de dados no DW.

Machado (2013, p.59) complementa que a granularidade de dados remete ao “nível de sumarização dos elementos e de detalhe disponíveis nos dados. Quanto mais detalhe existir, mais baixo será o nível de granularidade. Quanto menos detalhe existir, mais alto será o nível de granularidade”.

Turban e outros (2009) deixam claro que quanto menor o nível de granularidade, maior é a quantidade de dados armazenados no DW, conseqüentemente, afetando o desempenho das consultas deixando o tempo de resposta mais longo. Por isso, a importância

de identificar o nível necessário de granularidade na definição de um escopo de um projeto de DW.

Segundo Inmon (2005) granularidade sempre será o maior problema de modelagem, dentro de uma mesma companhia, áreas diferentes podem solicitar os dados em níveis diferentes, por exemplo, o *marketing* pode querer ver informações de vendas por mês, enquanto o setor financeiro deseja ver vendas por semana. Na percepção dos autores a maior dificuldade é atender as necessidades de todas as áreas.

Quanto maior o nível de granularidade dos dados, maior a perda na capacidade de responder a consultas detalhadas usando os dados. [...] com um nível alto de granularidade, os dados se tornam compactos e podem fornecer uma resposta rápida. Desde que atenda a necessidade, as consultas em um alto nível de granularidade serão muito mais eficientes (INMON, 2005, p.41).

Machado (2013, p.60) reforça a importância de definir a granularidade em um projeto: “Quando a granularidade de um *Data Warehouse* ou de um *Data Mart* é estabelecida, o projeto tem o seu desenvolvimento com foco e irá fluir com mais tranquilidade. Enquanto não for visualizado o nível de granularidade, torna-se muito difícil a modelagem do *Data Mart*”.

Consequentemente, quanto maior o nível de granularidade dos dados, menor a possibilidade de responder a consultas detalhadas. Por exemplo, se é definido que a necessidade é ver dados mensalmente, em um primeiro momento, não é possível ver os dados semanalmente. Com um nível alto de granularidade, os dados se tornam compactos e podem fornecer uma resposta rápida. Desde que atenda a necessidade, as consultas em um alto nível de granularidade serão muito mais eficientes (INMON, 2005).

2.2.7 Staging Area

Staging area é uma camada intermediária entre a fonte de dados e do DW, nessa parte pode ser feita uma limpeza inicial dos dados (CECI, 2012), pode ser considerado uma área de construção para o *data warehouse*. Muitas vezes, dependendo do tipo de dados ou da necessidade, o banco de dados de *staging* não precisa ser baseado em tecnologia relacional.

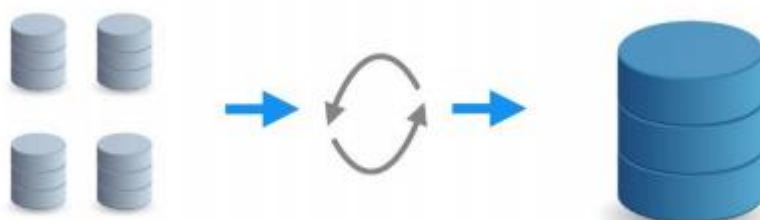
Para Kimball (1998), *staging area* pode ser interpretado como a área de preparo de um restaurante, onde os ingredientes crus são manuseados de forma a criar refeições refinadas.

Da mesma maneira que os cozinheiros de preparo não lidam com os consumidores, seja por organização, higiene, proteção, o banco de dados de *staging* não é ligado diretamente ao aplicativo final. A organização e padrões simples bem definidos nessa fase é essencial para seu bom funcionamento, para evitar que o processo possa se tornar caótico e confuso até mesmo para o desenvolvedor responsável.

2.3 ETL

ETL significa *Extract, Transform and Load*, em outras palavras, na sua tradução literal significa Extração, Transformação e Carga. Este processo é mostrado na figura 6

Figura 6 – Processo ETL



Fonte: Piton (2018, p. 242).

Conforme Kimball (2003) ETL é a fundação do DW. É uma etapa essencial que necessita de um bom planejamento e não é aconselhável apressar essa fase, pois quando bem projetada reforça a qualidade dos dados e padronização consistente. Esse processo também ajusta os dados para que fontes separadas possam ser usadas juntas, entregando dados em um formato para que o desenvolvedor possa criar por exemplo aplicativos, relatórios, infográficos e gráficos para que o usuário final tenha informações para tomar decisões.

Ainda segundo as opiniões de Kimball (2013), ETL é uma atividade *backend* e não vista pelo usuário final, porém pode consumir facilmente mais de 70% dos recursos destinados à implementação e manutenção do DW. O ETL é muito mais que um meio de pegar os dados do banco de origem para o *data warehouse*:

- a) Remove erros e corrige dados não padronizados ou não existentes na estrutura.

- b) Provê documentação das medidas.
- c) Captura o fluxo de dados.
- d) Ajusta os dados de múltiplas fontes de origem.
- e) Estrutura os dados para serem utilizados em ferramentas diversas.

No passo da extração é onde se obtém os dados do sistema de origem ou de um outro DW, um dos principais objetivos desta fase é liberar o sistema de origem do processamento dos próximos passos. Na transformação são criadas medidas, limpeza de dados, seleção de colunas que serão utilizadas, unificação dos dados no contexto desejado, mesclagem e controle dos dados, como por exemplo geração de chaves primárias. A terceira e última fase é a parte na qual carrega os dados tratados em um *database* ou em um *staging area* com a possibilidade de utilizar técnicas como carregamento em lotes ou carregamento de forma distribuída (PITON, 2018).

2.4 OLAP

Online Analytical Processing, (OLAP) é um grupo de ferramentas que consegue explorar os dados do DW, ou seja, tem a capacidade de analisar dados em uma arquitetura multidimensional em vez de tabelas padrões (BARBIERI, 2011).

Ainda conforme as palavras de Machado (2013, p.85):

[...] o usuário tem uma visão dos dados de um Data Warehouse, podendo efetuar operações ditas bárias como slice and dice, que é uma forma de mudança das dimensões a serem visualizadas, drill down e roll up, como se denomina a navegação entre os níveis de detalhamento dos dados do Data Warehouse.

Com essa tecnologia é possível manipular e analisar um grande volume de dados em tempo real, tendo uma visão de múltiplas perspectivas, pode ser considerado uma tela em branco, para que o usuário possa construir gráficos, filtros, medidas, colunas e outros para enxergar mais rapidamente possível a informação desejada (PITON, 2018).

Conforme Kimball (2004), os servidores OLAP possuem duas funções principais

- a) Desempenho de *query*: ela gerencia automaticamente, agregando e criando indexes, melhorando a velocidade de consulta e minimizando o tempo de criação de processos de *tunning*.
- b) Riqueza analítica: utiliza linguagens muito bem adaptadas e relativamente rápidas para análise complexa de dados, diferente do SQL. Além disso, possui mecanismos que armazenam cálculos complexos e configurações de segurança.

OLAP um grande aliado no processo de tomada a decisão, ajudando o analista de negócios a criar visões úteis e complexas em pouco tempo e de forma simples, conseguindo identificar problemas, *gaps* e padrões em minutos, sem precisar fazer requisições para uma equipe de tecnologia da informação para cada necessidade de análise e busca. (PITON, 2018).

2.4.1 Operações

Antes de falar sobre operações OLAP, é preciso entender o que é um cubo dimensional, definido pela Microsoft (2016a) como:

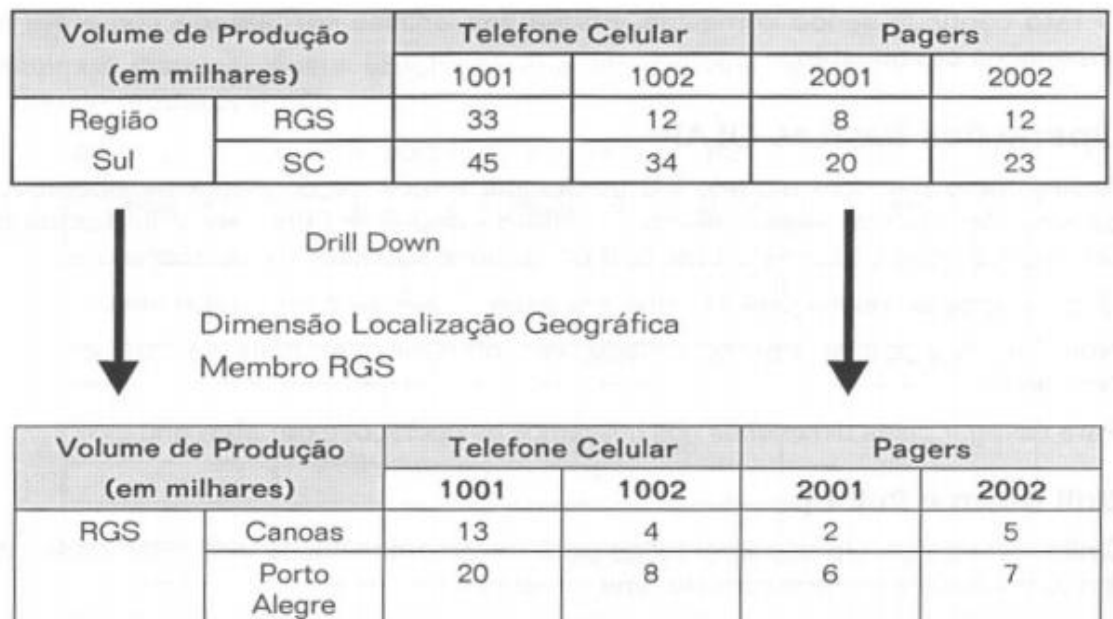
Um cubo OLAP é uma estrutura de dados que supera as limitações dos bancos de dados relacionais, proporcionando rápida análise de dados. Os cubos podem exibir e somar grandes quantidades de dados enquanto fornecem aos usuários acesso pesquisável a quaisquer pontos de dados. Dessa forma, os dados podem ser acumulados, segmentados e separados, conforme necessário, para controlar a mais ampla variedade de perguntas relevantes para a área de interesse de um usuário.

Machado (2013) acredita que existem quatro tipos de operações utilizadas em OLAP para analisar dados: *Drill Down* e *Drill Up*, *Drill Across*, *Drill Throught* e *Slice and Dice*.

2.4.1.1 Drill Down e Drill Up

Operações definidas por Machado (2013, p.86) como “operações para movimentar a visão dos dados ao longo dos níveis hierárquicos de uma dimensão”. O mesmo autor explica que as operações de *drill* permitem que a informação seja visualizada pelo usuário do nível mais alto até o mais detalhado, como também a possibilidade de partir da informação mais detalhada e sumariza-la.

Figura 7 – Drill Down




Fonte: Machado (2013, p.87).


Ceci (2012, p.137) exemplifica e explica que com o *drill down*, é possível “ver as informações agrupadas por ano, mês, semana, dia até chegar no nível de operação”.

Figura 8 – Drill Up

Volume de Produção (em milhares)		1999			
		Trim. 1	Trim. 2	Trim. 3	Trim. 4
Região Sul	RGS	78	67	22	56
	SC	90	67	88	99



Roll Up
Dimensão Tempo



Volume de Produção (em milhares)		Trimestre 1		
		Janeiro	Fevereiro	Março
Região Sul	RGS	30	26	22
	SC	28	30	32

Fonte: Machado (2013, p.86).

Machado (2013) explica, de acordo com a figura 7, que o *drill down* possibilita aumentar o nível de detalhamento da informação, ou seja, diminuir a granularidade. *Drill up*, também conhecido como *roll up*, é a operação contrária, aumentando a granularidade, é possível partir do nível mais detalhado para o mais sumarizado, conforme figura 8.

2.4.1.2 Drill Across

Outra operação citada por Machado (2013, p.88) é o *drill across* que “ocorre quando o usuário pula um nível intermediário dentro de uma mesma dimensão”. Com isso é possível ignorar a camada de um grão maior para um grão menor.

Sell (2006, p.37) complementa com um exemplo de *drill across*: “o nível da análise é alterado direto de ano para mês dentro de uma dimensão tempo, quando esta é composta de ano, semestre e mês”.

2.4.1.3 Drill Throught

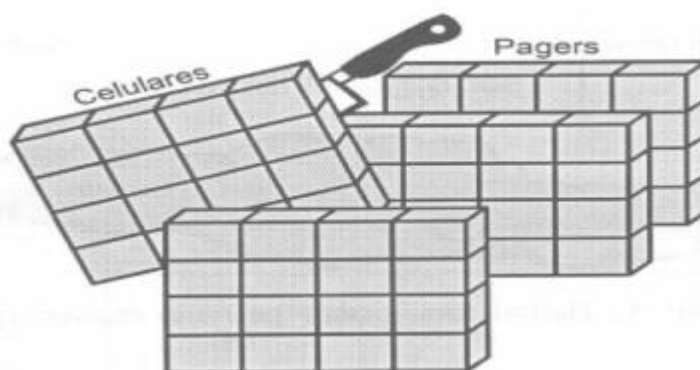
Conforme Sell e Machado (2006, 2013) *drill throught* é quando o usuário passa a visualizar a informação de uma dimensão para outra. Ambos autores trazem o exemplo de quando o usuário está olhando dados pela dimensão tempo e em seguida passa a analisar os mesmos dados pela dimensão de região.

2.4.1.4 Slice and Dice

Machado (2013, p.89) explica que “*Slice and Dice* significa em uma forma simplista a redução do escopo dos dados em análise, além de mudar a ordem das dimensões, mudando desta forma a orientação segundo a qual os dados são visualizados”.

Machado (2013) reforça que *slice and dice* é uma das características mais importantes de ferramentas OLAP, essas operações permitem a análise da informação sobre visões que anteriormente não existiam e/ou nunca havia se pensado. Tornou-se possível “analisar as informações de diferentes primas limitados somente pela nossa imaginação” (MACHADO, 2013, p.89). Tudo isso devido a criação do DW e utilizações de ferramentas OLAP.

Figura 9 – Slice



Fonte: Machado (2013, p.90).

Segundo Sell (2006) e conforme a figura 9, *Slice* opera como um filtro, exibindo apenas parte dos dados uma dimensão de acordo com o que é selecionado, por exemplo, visualizar somente dados do ano de 2018 em uma perspectiva. Já o *dice* como a inversão de perspectiva da visão multidimensional, Sell (2006, p.37) exemplifica: “A análise é alterada de região (linha) por ano (coluna) para ano (linha) por região (coluna) ”.

2.4.2 Medidas

Segundo Machado (2013, p.81): “Medidas são atributos numéricos que representam um fato, a performance de um indicador de negócios relativo às dimensões que participam desse fato”. O Mesmo autor exemplifica medidas: “valor de vendas em reais, número de produtos vendidos, quantidade em estoque, custo de venda, entre outros”.

No que diz respeito a objetivo, Sell (2006, p.31) afirma que as medidas “representam o desempenho de um indicador de negócios relativo às dimensões que participam de um fato”.

Medidas, de acordo com Microsoft (2016b) são utilizadas para o cálculo e exibição de indicadores, como por exemplo, valor total de vendas por cliente, calculo que envolve a função de agregação de soma, além dessa, outras funções de agregação mais comum incluem mínimo, máximo e contagem.

2.5 JUDICIÁRIO

Conforme exposto por CNJ e a Constituição Federal brasileira (2015, 1988) o Poder Judiciário é dividido em vários órgãos diferentes para solucionar problemas e conflitos da sociedade, garantindo os direitos dos cidadãos. A área civil busca solucionar conflitos que surgem entre pessoas e empresas. A penal impõe pena naqueles que cometem crimes. Na parte

federal, julgam casos que sejam de interesse do país, além dessas, existe também a área trabalhista, militar e eleitoral, cada um com sua jurisdição.

O Poder Judiciário tem como cúpula o Supremo Tribunal Federal (STF), sendo o CNJ o responsável pelo controle da atuação administrativa e financeira. Os órgãos do judiciário normalmente são classificados em relação ao número de julgadores, quanto a matéria de especialização e de ponto de vista federativo (CNJ, 2015).

2.5.1 Tribunais

De acordo com CNJ (2010): “os Tribunais de Justiça (TJs) são responsáveis por revisar os casos já analisados pelos juízes singulares de primeira instância. São 27 TJs, um em cada unidade da Federação, cuja competência é julgar recursos das decisões dos juízes de primeiro grau”.

2.5.2 CNJ

Conselho Nacional de Justiça (CNJ) é uma instituição pública com o objetivo de melhorar o trabalho do sistema judiciário brasileiro como um todo, uma das funções principais dela é no que diz respeito à transparência e ao controle de toda a organização. (CNJ, [2009?])

A missão do conselho é de “desenvolver políticas judiciárias que promovam e efetividade e a unidade do Poder Judiciário, orientadas para os valores de justiça e paz social”.

Conforme Folha (2012), o CNJ foi criado em 2004 pela Emenda Constitucional 45, conhecida como ‘Reforma do Judiciário’, para exercer a atribuição de um controle externo do Judiciário, tendo a função do planejamento estratégico baseado em dados e gestão administrativas dos tribunais em operação e um controle disciplinar e correccional das atividades dos magistrados.

Conforme o CNJ ([2009?]), a organização tem a responsabilidade de:

- a) Zelar pela autonomia do Poder Judiciário e pelo cumprimento do Estatuto da Magistratura.
- b) Definir planejamento estratégico, planos de metas e programas de avaliação institucional.
- c) Receber reclamações, petições eletrônicas e representações contra membros ou órgãos do Judiciário
- d) Julgar processos disciplinares, assegurando ampla defesa.
- e) Definir melhores práticas e celeridade, elaborando e publicando semestralmente relatórios estatísticos sobre movimentação processual e outros indicadores pertinentes à atividade jurisdicional em todo país.

Ainda de acordo com a Folha (2012), existiu dois grandes motivos para essa reforma, o primeiro é pelo motivo de que o judiciário brasileiro é composto por vários tribunais diferentes, cada um deles com sua autonomia total, com o CNJ foi possível criar padrões de qualidades e diretrizes a serem seguidos. O segundo foi pela crítica ao corporativismo da magistratura e a necessidade de existir algum controle sobre os três poderes, com a reforma o CNJ possui o poder de correição e controle disciplinar, seja investigando, corrigindo ou punindo os responsáveis.

2.5.2.1 Relatório Justiça em Números

A resolução n. 76 de 12 de maio de 2009 determina que os Tribunais devem dispor os dados e de fácil acesso para divulgação de dados estatístico relativos à atuação administrativa e jurisdicional, inclusive produtividade dos magistrados (ROVER, 2016).

De acordo com CNJ (2017): “os indicadores e as variáveis calculados são fundamentados na Resolução CNJ n. 76/2009, que, em seus anexos, traz o detalhamento de glossários e fórmulas que norteiam o Sistema de Estatísticas do Poder Judiciário (SIESPJ)”.

Produzido anualmente pelo Departamento de Pesquisas Judiciárias (DPJ), unidade responsável pelo recebimento e pela sistematização das estatísticas judiciárias nacionais, o relatório de 2017 traz uma nova organização, com os capítulos estruturados a partir dos indicadores, e não mais em relação a cada um dos segmentos de Justiça. Dessa forma, todos os ramos de Justiça são apresentados em conjunto,

possibilitando uma análise mais completa do desempenho do Poder Judiciário, o que permite leitura mais apropriada das características de cada um dos tribunais e das diferenças entre os ramos de Justiça (CNJ, 2017, p.9).

O mais recente relatório contém as informações entre os anos de 2009 a 2016, com informações dos órgãos do Judiciário brasileiro, excluindo o Supremo Tribunal Federal e o Conselho Nacional de Justiça (CNJ, 2017).

CNJ (2017) explica que o SIESPJ agrupa os indicadores estatísticos necessários do judiciário em principais categorias:

- a) Insumos, dotações e graus de utilização: informado anualmente, são compostos por dados de receitas, despesas e estrutura.
- b) Litigiosidade: informados semestralmente, contém dados de carga de trabalho, taxa de congestionamento, produtividade e recorribilidade.
- c) Escolaridade: informados anualmente, provém dados de acesso à justiça e tempo do processo.

CNJ (2017) deixa claro que os tribunais de justiça são obrigados a fornecer dados dos indicadores definidos pelo SIESPJ, além disso, são totalmente responsáveis pelas informações apresentadas, cabendo aos tribunais definir um responsável para gerar, conferir e transmitir dados estatísticos.

3 METODOLOGIA

Nesse capítulo é abordado qual método de pesquisa foi escolhido para apresentação desse trabalho, além das etapas metodológicas e delimitações.

3.1 CARACTERIZAÇÃO DO TIPO DE PESQUISA

Conforme Silva e Menezes (2005, p. 19), de uma forma simplificada, pesquisa significa "procurar respostas para indagações propostas", ou seja, um conjunto de ações e propostas com o objetivo de alcançar solução de problemas, porém para isso é necessário usar passos racionais e sistemáticos.

É possível classificar a pesquisa em duas formas: básica e aplicada. A metodologia de pesquisa adotada neste trabalho é pesquisa aplicada, pois tem como objetivo trazer o conhecimento para o público, dando acessibilidade e visibilidade destes dados judiciais. De acordo com Silva e Menezes (2005, p.20), uma pesquisa aplicada tem como objetivo “gerar conhecimentos para aplicação prática e dirigidos à solução de problemas específicos. Envolve verdades e interesses locais”.

Quando se trata da abordagem do problema em si, a pesquisa pode ser quantitativa ou qualitativa. Neste trabalho é utilizada a abordagem qualitativa, pois conforme Gil, Silva e Menezes (2002, 2005):

- a) Pesquisa quantitativa - considera que tudo pode ser traduzido em números, seja opiniões ou informações, as categorias são pré-estabelecidas, então o trabalho analítico é simplificado, ela requer uso de técnicas estatísticas.
- b) Pesquisa qualitativa - o conjunto de categoria é modificado e refeito várias vezes no processo com o objetivo de obter resultados mais significativos e abrangentes, acredita que muitos assuntos não podem ser traduzidos em

números, a dinâmica com o mundo real traz subjetividade, o processo e o seu significado são os focos principais.

No que diz respeito a objetivos, este trabalho se encaixa em uma pesquisa exploratória. Gil, Silva e Menezes (2002, 2005) classifica e explica objetivos em pesquisas de três grupos:

- a) Exploratória: trazer mais conhecimento sobre o problema em questão, tornando possível torna-lo mais claro e com possibilidade de construir hipóteses, seu principal objetivo é aprimoramento de ideias ou descoberta de intuições.
- b) Descritiva: descrever características de determinada população, fenômeno ou estabelecimento de relações entre variáveis, atrelada a estatística, tem como por exemplo, estudar as características de um determinado grupo.
- c) Explicativa: quando se deseja explicar as causas de algum fenômeno, aprofunda o conhecimento da realidade, justificando cada coisa, é considerado complexo e delicado, pois o risco de cometer erros aumenta consideravelmente.

Do ponto de vista de procedimentos técnicos, Gil, Silva e Menezes (2002, 2005) classificam as pesquisas em:

- a) Pesquisa Bibliográfica.
- b) Pesquisa Documental.
- c) Pesquisa Experimental.
- d) Levantamento.
- e) Estudo de caso.
- f) Pesquisa *Expost-Facto*.
- g) Pesquisa Ação.
- h) Pesquisa Participante.

De acordo com a classificação de Gil, Silva e Menezes (2002, 2005), este trabalho se trata de uma pesquisa bibliográfica, pois é realizada a partir de material já publicado, como por exemplo: livros e artigos.

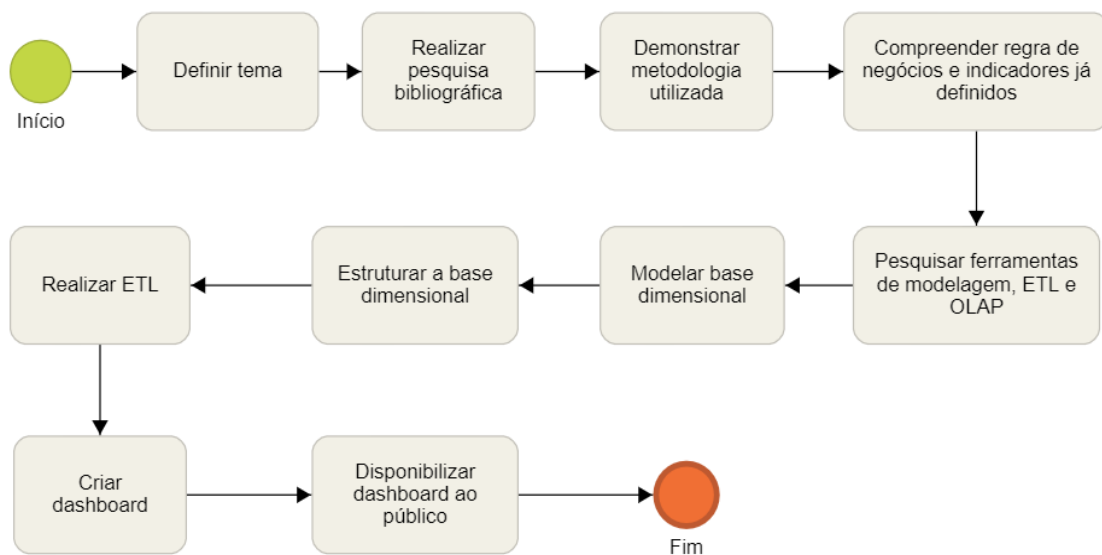
3.2 ETAPAS METODOLÓGICAS

Este trabalho é dividido em diversas etapas, segundo as atividades descritas abaixo:

- f) No primeiro capítulo foi elaborada a definição do tema, a problemática, declaração dos objetivos gerais e específicos e a justificativa.
- g) A segunda etapa é constituída pela pesquisa bibliográfica, onde é feito a base do conceito do trabalho, sustentando com teorias de especialistas e principais autores da área.
- h) Na terceira etapa é demonstrada a metodologia utilizada no trabalho, descrevendo as etapas e delimitações e cronograma para a contextualização do mesmo.
- i) O quarto capítulo é o desenvolvimento e a implementação em si, de acordo com a proposta e os objetivos criados no primeiro capítulo.

Na figura 10 é possível visualizar as atividades no fluxograma representando as etapas citadas.

Figura 10 – Etapas Metodológicas



Fonte: Elaborado pelos autores.

O primeiro passo do desenvolvimento será compreender regra de negócios com especialistas na área e indicadores previamente já definidos pelo SIESPJ e CNJ, o segundo passo é uma etapa na qual será realizada uma pesquisa por melhores ferramentas de modelagem,

ETL e OLAP que atenda aos requisitos mínimos, no terceiro passo será modelado a estrutura dimensional, no quarto passo será estruturado o modelo dimensional definido no passo anterior, o quinto passo se dará pelo desenvolvimento de ETL, o sexto passo será a criação de visões de indicadores em *dashboards*, o sétimo e último passo é a disponibilização desse *dashboard* ao público.

3.3 DELIMITAÇÕES

Este trabalho tem como objetivo desenvolver uma solução de BI com os dados abertos disponibilizados pelo CNJ, dado as seguintes limitações:

- a) Não serão feitas atualizações automáticas dos dados.
- b) Os dados utilizados serão só os disponibilizados pelo CNJ.
- c) Serão desenvolvidos somente os requisitos necessários para validar o objetivo proposto.
- d) Não será feita a interpretação dos dados obtidos pelos autores.

4 MODELAGEM

Neste capítulo é apresentada a modelagem dimensional do projeto de DW. Começando pelas regras de negócio, requisitos funcionais e não funcionais, apresentado as regras de negócio e alguns indicadores utilizado pelo ‘Relatório Justiça em Números’.

4.1 REQUISITOS

Segundo Sommerville (2011), os requisitos de um sistema é um conjunto de descrições muito importante para todo o processo, seja colocando restrições no funcionamento, ou especificando o que o sistema deverá fazer. O requisito na área da computação não possui um padrão bem definido, em muitos casos pode ser apenas uma especificação em alto nível, em outros é feito com detalhamento. Essas diferenças nas estruturas e níveis de detalhamento podem ser explicadas pela diversidade das necessidades da empresa/cliente.

Ainda conforme Sommerville (2011), os requisitos de softwares são normalmente classificados em dois grupos, o requisito funcional e o não funcional. Mesmo teoricamente possuindo características distintas entre si, a diferenciação não é tão clara como sugere as definições, um dos motivos disso é que os requisitos não são independentes entre si e podem gerar ou restringir outros requisitos.

4.1.1 Perguntas estratégicas

Quando se desenvolve uma solução utilizando *business intelligence* normalmente existe a necessidade de responder questões sobre o negócio ou questões específicas que o cliente

tenha. Como explicado no capítulo dois, as principais perguntas que costumeiramente o BI responde são:

- a) “O que aconteceu?”
- b) “Quantas vezes isso aconteceu?”
- c) “Quando isso ocorreu?”
- d) “Onde ocorreu?”
- e) “Quem fez isso?”

Após isso, é preciso entender quem é o público alvo do experimento. Neste trabalho o grupo alvo é bem variado, por se tratar de dados de justiça disponíveis para todos os cidadãos, porém o que essas pessoas possuem em comum é o interesse e necessidade de informação sobre justiça de forma simples, rápida, intuitiva e comparativa.

Com isso em mente e após análise dos dados disponibilizados pelo Conselho Nacional de Justiça, foi escolhida algumas perguntas estratégicas que o relatório deverá responder.

- a) Quais são os indicadores?
- b) Qual é o histórico do indicador selecionado?
- c) Qual o porte atual do tribunal e seu histórico?
- d) Qual é a média do indicador selecionado?
- e) Comparativo do tribunal selecionado com outros tribunais do mesmo porte.
- f) Qual a média de todos os outros tribunais desse indicador?
- g) Qual a média por porte do indicador selecionado?
- h) Qual a categorização e subcategorização de um determinado indicador?
- i) Quais são os tipos de tribunais disponíveis para análise?

Com essas perguntas chaves definidas, é possível criar a regra de negócio, modelagem e o relatório, como mostrado nas próximas seções.

4.1.2 Regras de negócio

De acordo com Vazquez (2016) regras do negócio normalmente são as leis que ditam o negócio e dizem como deve funcionar seus processos. É a forma o qual o comportamento do negócio é seguido mediante a itens como: legislação, regulamentações e padrões de mercado.

O Quadro 3, detalha as regras de negócio levantadas nesse projeto:

Quadro 3 - Regras de Negócio

Regra de Negócio	Descrição
RN01 - Dimensão dTempo	O sistema deve conter uma tabela de tempo de acordo com os períodos da(s) tabela(s) fato. A tabela deverá conter os campos: a) skData b) nuAno c) flEleicao
RN02 - Dimensão dTribunal	O sistema deve conter uma tabela dos tribunais com os dados disponibilizados pelo CNJ. A tabela deverá conter os campos: a) skTribunal b) cdTribunal c) nmTribunal d) nmUf e) nmRegiao f) nmJustica g) nuOrdem
RN03 - Dimensão dIndicador	O sistema deve conter uma tabela que contenha todos os indicadores disponibilizados pelo CNJ. A tabela deverá conter os campos: a) skIndicador b) cdIndicador c) nmIndicador d) deFormula

RN04 - Fato Indicadores	<p>O Sistema deverá conter a métrica do valor do indicador e deverá conter os campos:</p> <ul style="list-style-type: none"> a) skData b) skTribunal c) skIndicador <p>A tabela deverá fazer relacionamento com as seguintes dimensões:</p> <ul style="list-style-type: none"> a) dTempo b) dTribunal c) dIndicador
RN05 - Extração de Dados	O Sistema deverá apenas realizar a extração de dados disponibilizados pelo CNJ.
RN06 - Transformação de Dados	O sistema deverá transformar os dados para que se adequem a estrutura do modelo estrela previamente definida.
RN07 - Carga de Dados	O sistema deverá inserir os dados estruturados no banco de dados após a etapa de transformação.
RN08 - Modelo ER	O sistema deverá disponibilizar o modelo ER.

Fonte: Elaborado pelos autores.

Na próxima subseção, são apresentados os requisitos funcionais.

4.1.3 Requisitos Funcionais

Segundo as palavras de Sommerville (2011), o requisito funcional deve dizer o que o sistema deve fazer, essas informações são dependentes e alinhadas com o tipo do *software*, responsáveis, estratégias da organização, dos usuários.

Citando Sommerville (2011, p. 60) “consistência significa que os requisitos não devem ter significados contraditórios [...]. Uma razão para isso é que [...] elaborar especificações para sistemas complexos é fácil cometer erros e omissões”. Os requisitos

funcionais em princípio devem ser completos e consistentes, porém, como dito anteriormente, pode variar de requisitos bem específicos até requisitos gerais e abrangentes.

O Quadro 4, detalha os requisitos funcionais levantados nesse projeto:

Quadro 4 - Requisitos Funcionais

Regra de Negócio	Descrição
RF01 - Inserção de dados	O sistema não deve permitir inserção de dados pelo usuário final.
RF02 - Remoção de dados	O sistema não deve permitir a remoção de dados pelo usuário final.
RF03 - Exportação de dados	O sistema deve permitir a exportação de até 50.000 linhas de dados.
RF04 - Consultas OLAP	O sistema deve permitir realizar funcionalidades de acesso, visualização e análise de dados

Fonte: Elaborado pelos autores.

Na próxima subseção, são apresentados os requisitos não funcionais.

4.1.4 Requisitos não Funcionais

Os requisitos não funcionais não estão ligados de forma direta com o *software* para os usuários finais como os requisitos funcionais. Alguns exemplos de requisitos não funcionais são de proteção, desempenho e disponibilidade. Esses requisitos são muito importantes e até mais críticos que o próprio requisito funcional, pois uma falha pode inutilizar o sistema como um todo (SOMMERVILLE, 2011).

Uma informação muito importante de acordo com Sommerville (2011) nos requisitos funcionais é a quantificação, quando possível do objetivo, pois dessa forma podem ser testados mais facilmente. Com frequência esses requisitos interagem com os requisitos funcionais, e como explicado, o trabalho de separar em duas classes distintas pode ser consideravelmente complexo.

O Quadro 5, detalha os requisitos não funcionais levantados nesse projeto:

Quadro 5 - Requisitos não Funcionais

Regra de Negócio	Descrição
RNF01 - Banco de Dados	O sistema deve utilizar o banco de dados SQL SERVER 2017 Developer.
RNF02 - ETL	O sistema deve utilizar o Pentaho Data Integration Community 7.1 para etapas de ETL.
RNF03 - Visualização de Dados	O sistema deve utilizar o Microsoft Power BI versão 2.62.5222.761 como ferramenta de visualização de dados.
RNF04 - Modelagem Entidade Relacional	O sistema deve utilizar o Power Architect v1.0.8 Community Edition para visualização do modelo entidade relacional (ER).
RNF05 - Modelagem Entidade Relacional	O sistema deve permitir a exportação de até 50.000 linhas de dados em formato .csv e .xlsx.

Fonte: Elaborado pelos autores.

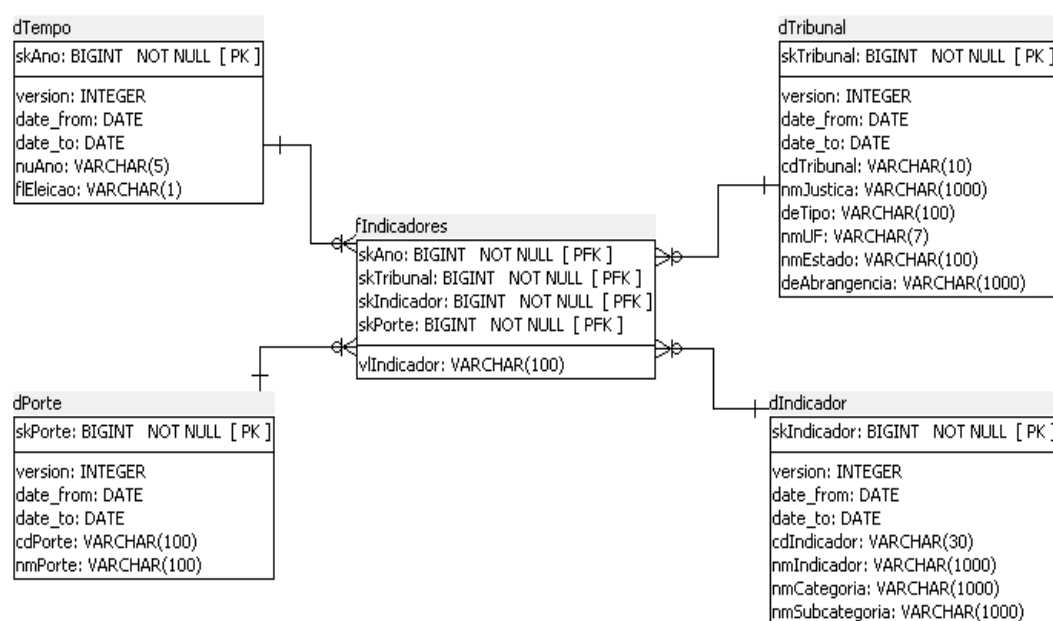
Na próxima sessão, é apresentado o modelo dimensional do projeto.

4.2 MODELAGEM DIMENSIONAL

Para esse projeto de DW, foi optado pelo esquema estrela pela justificativa de que encontraria melhor performance no processamento de dados.

O modelo lógico possui uma tabela fato fIndicadores e três tabelas dimensões: dTempo, dTribunal e dIndicador, conforme a figura 11.

Figura 11 – Modelagem Dimensional



Fonte: Elaborado pelos autores.

Nas próximas subseções, são apresentados as tabelas dimensões e fato que compõem o modelo dimensional.

4.2.1 Dimensão Tempo

A dimensão tempo, representada pela tabela dTempo, conforme figura 11, armazena dados de ano (coluna *nuAno*) e se houve eleições naquele ano (coluna *flEleicao*).

4.2.1 Dimensão Tribunal

A dimensão tribunal, representada pela tabela *dTribunal*, conforme figura 11, armazena dados de nome do tribunal (coluna *nmTribunal*), o código na fonte de dados (coluna *cdTribunal*), o tipo de justiça (coluna *deTipo*), nome da unidade federativa (coluna *nmUF*), nome da região (coluna *nmRegiao*), o nome do estado (coluna *nmEstado*), a descrição da abrangência (coluna *deAbrangencia*) e tipo de justiça (coluna *nmJustica*).

4.2.2 Dimensão Indicador

A dimensão indicador, representada pela tabela *dIndicador*, conforme figura 11, armazena dados de indicadores disponibilizados pelo CNJ, armazena dados de código do indicador (coluna *cdIndicador*), nome do indicador (coluna *nmIndicador*), nome da categoria (coluna *nmCategoria*) e o nome da subcategoria (coluna *nmSubcategoria*).

4.2.3 Dimensão Porte

A dimensão porte, representada pela tabela *dPorte*, conforme figura 11, armazena dados de tamanho do porte do tribunal (coluna *nmPorte*) e o código na fonte de dados (coluna *cdPorte*).

4.2.4 Fato Indicadores

A tabela fato indicador, representada pela tabela fIndicador, conforme figura 11, faz ligação com todas as dimensões do modelo e possui a medida do valor do Indicador (coluna *vlIndicador*) e como explicado no capítulo dois, o *surrogate key* para a relação com as tabelas dimensões.

5 DESENVOLVIMENTO

Neste capítulo é apresentado as ferramentas utilizadas, o processo de desenvolvimento e resultados obtidos.

5.1 FERRAMENTAS UTILIZADAS

Nesta seção irá ser exposto as ferramentas utilizadas durante o processo de desenvolvimento, seja para armazenamento, transformação, extração, carregamentos, modelagem e outros.

5.1.1 Microsoft Power BI

Power BI é um serviço de *Business Intelligence* da *Microsoft*, oferecendo além de visualizações interativas, conectores e aplicativos para utilizar de forma coerente até fonte de dados não relacionados. Os principais fatores positivos são o baixo investimento inicial, escalável, ampla documentação, disponível em vários idiomas e ser *self-service*, ou seja, usuários finais podem criar as suas análises e relatórios sem depender de um profissional técnico.

Figura 12 – Quadrante Mágico da Gartner – Analytics and Business Intelligence Plataform



Fonte: Gartner (2018).

Conforme mostrado na figura 12, em 2018, o Power BI é um dos destaques do quadrante mágico de “*Analytics and Business Intelligence Plataforms*” da Gartner, demonstrando a evolução constante da ferramenta.

5.1.2 Microsoft SQL Server

SQL Server é um gerenciador de banco de dados relacional desenvolvido pela Microsoft, seu principal objetivo, como a maioria dos bancos de dados é armazenar e recuperar os dados que forem solicitados.

Suas duas principais linguagens de consultas são o *Transact-SQL* e *SQL*.

5.1.3 SQL Server Management Studio

SQL Server Management Studio (SSMS) é uma ferramenta para consultar, gerenciar, criar os bancos de dados e *data warehouses*, seja na nuvem ou no computador local. Ele inclui editor de *scripts* e utensílios gráficos.

Permite o usuário navegar, selecionar e utilizar qualquer objeto dentro do servidor, como se fosse um *object explorer*, tudo isso baseado no *shell* do Visual Studio, possui também compatibilidade retroativa e é independente da versão da *engine* do banco de dados SQL Server, sendo de simples utilização, baixa curva de aprendizado e fácil gerenciamento de *drivers* nativos.

5.1.4 Pentaho Data Integration Kettle

Pentaho Data Integration (PDI), mais conhecido como Kettle, faz parte da suíte da Pentaho. Nesse trabalho é utilizado a versão *Community*, que é a versão gratuita, porém possui os mesmos recursos da versão paga. Kettle provê recursos de extração, transformação e carregamento, facilitando a limpeza, padronização e armazenamento para os usuários e tecnologias que irão utilizar os dados tratados.

Pentaho Data Integration é utilizado no mercado para várias finalidades como por exemplo:

- a) Migração de dados entre diferentes bancos de dados.
- b) Limpeza de dados simples ou complexas.
- c) Integração de várias fontes de dados em um repositório único.
- d) Criação de uma forma simples e intuitiva de *surrogate key*, *snapshots* e dimensões.

PDI é uma ferramenta escrita em Java, *open source*, além disso, possui inúmeros *plug-ins* criados por usuários diversos, documentação atualizada e comunidade ativa. Kettle é formado basicamente por três artefatos Spoon, Kitchen e Pan.

- a) Spoon – Interface gráfica da aplicação desktop, onde é possível criar transformações e Jobs.
- b) Pan – Programa que executa as transformações, normalmente agendada de modo automático por batch.
- c) Kitchen – Programa que executa os Jobs, também pode ser agendada em intervalos regulares por programas de gerenciamento, como o Pan.

A transformação nada mais é que uma sequência lógica de tarefas, um fluxo de dados, cada passo é chamado de *step* e cada ligação entre elas é chamada de *hop*, porém cada passo não é executado de forma sequencial e sim de modo paralelo. O *Job*, diferente da transformação, é um fluxo de operação feita de forma sequencial, é possível combinar por exemplo, transformações em uma sequência predeterminada, além disso, é possível criar *Jobs* de armazenamento de *logs* ou mandar um *e-mail* para o responsável se ocorrer algum problema na transformação.

5.1.5 Microsoft Excel

Excel é um *software* da Microsoft para criação e edição de planilhas eletrônicas, com a opção de utilizar a linguagem *Visual Basic for Applications* (VBA). Neste trabalho o Excel

é utilizado para validação de dados, comparativos de tabelas, filtros simples e categorização manual de hierarquias.

5.1.6 SQL Power Architect

SQL Power Architect é um *software* da Best of BI criada por designers de DW e possui muitos recursos exclusivos voltados especificamente para o arquiteto de DW. Ele permite que os usuários façam engenharia reversa de bancos de dados existentes, executem a criação de perfis de dados em bancos de dados de origem e gerem automaticamente metadados ETL.

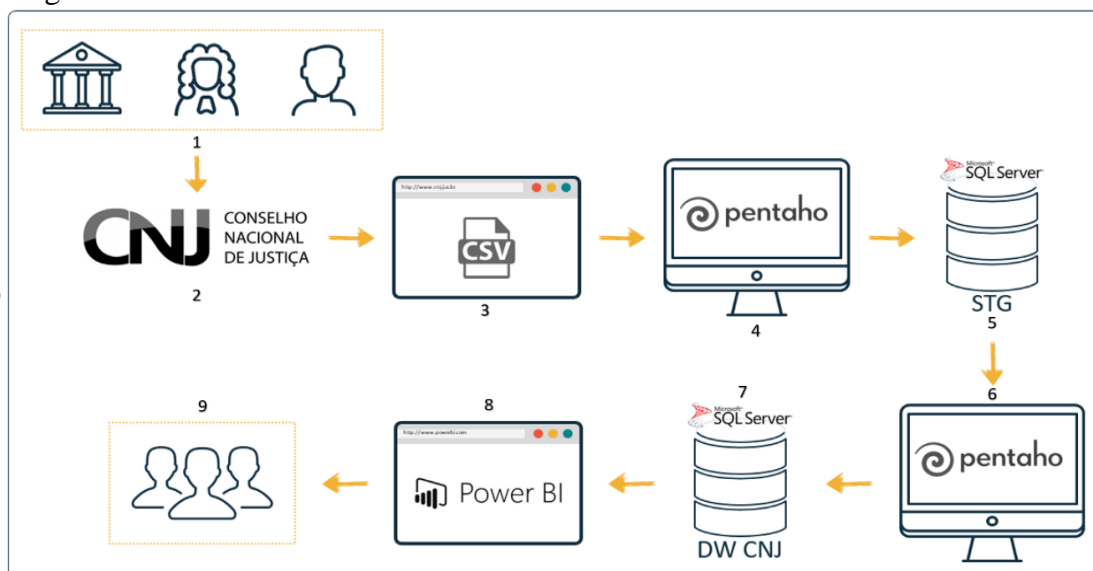
5.1.7 DAX

Data Analysis Expressions (DAX) é a linguagem utilizada pelo Microsoft Power BI para a criação de colunas e medidas personalizadas. Nela consta funções e operadores que podem ser combinadas para criar fórmulas simples ou complexas com o objetivo de fornecer uma melhor análise de dados para o usuário.

5.2 EXPERIMENTO

De acordo com a problemática e conforme o objetivo, ambos descritos no capítulo 1, este trabalho consiste na criação de uma solução de BI com os indicadores disponibilizados pelo CNJ. A figura 13 ilustra a solução completa e no quadro 6, as explicações de cada etapa do fluxo.

Figura 13 – Processo Desenvolvimento



Fonte: Elaborado pelos autores.

Quadro 6 – Processo de Desenvolvimento

Passo	Descrição
1	Os tribunais preparam o envio dos dados referentes aos indicadores definidos pelo SIESPJ.
2	O CNJ recebe e trata os dados enviados pelos tribunais.
3	Os dados são disponibilizados ao público no portal do CNJ em formato .csv para <i>download</i> .
4	Os dados passam por um tratamento prévio e necessário no Pentaho Data Intagracion.
5	Os dados são enviados para uma base SQL Server no banco STG.
6	Com os dados disponíveis no Banco STG, utiliza-se o Pentaho Data Intagracion para o início do processo de modelagem dimensional.
7	Os dados são enviados para uma base SQL Server no banco DWCNJ.
8	São criados relatórios utilizando Power BI com os dados disponíveis no banco DWCNJ.
9	Usuários finais consomem e utilizam o relatório criado.

Fonte: Elaborado pelos autores.

Na próxima subseção, são apresentadas informações sobre os arquivos utilizados neste trabalho.

5.2.1 Arquivos

Nessa etapa foi realizada o *download* da base de dados do Justiça em Números disponibilizada pelo CNJ e Instituto Brasileiro de Geografia e Estatística (IBGE). Os arquivos disponibilizados foram renomeados e descritos conforme o quadro 7.

Quadro 7 – Arquivos Fonte

Nome Origem	Novo Arquivo	Fonte	Descrição
Variaveis_27-Ago-2018.csv	Variaveis.csv	CNJ	Possui os dados de nome e id dos indicadores
JN_27-Ago-2018.csv	Indicadores.csv	CNJ	Possui os dados de valor dos indicadores por dados de tribunal e ano.
05eab2c73cfa6de84c1c02013ab1fd79.xls	Documentacao.csv	CNJ	Possui dados de classificação dos indicadores da Justiça Militar Estadual.
8c3f0d00a7f254aad6173049da0ed37d.xls	Documentacao.csv	CNJ	Possui dados de classificação dos indicadores da Justiça Eleitoral.
b487a7b61f7c6e46c89003faa8e8d5c6.xls	Documentacao.csv	CNJ	Possui dados de classificação dos indicadores da Justiça Federal.
b33a6cb00c8c942487e61ac3e6fffbe3.xls	Documentacao.csv	CNJ	Possui dados de classificação dos indicadores da Justiça do Trabalho.
140817d734eed2beab5d7c8ef1eb270a.xls	Documentacao.csv	CNJ	Possui dados de classificação dos indicadores da Justiça Estadual.
8cbc236817f57eca573f25b5581fc9ab.xls	Documentacao.csv	CNJ	Possui dados de classificação dos indicadores da Justiça Militar da União.

2f89d07e46d1d4ce7 613e6d5c5244a73.xls	Documentacao.csv	CNJ	Possui dados de classificação dos indicadores do Tribunal Superior Eleitoral.
fb3f40c13f9ff05425 d23abc306dcc4b.xls	Documentacao.csv	CNJ	Possui dados de classificação dos indicadores do Tribunal Superior do Trabalho.
3fed704d3cceed47d 667919b4a51376d.xls	Documentacao.csv	CNJ	Possui dados de classificação dos indicadores do Superior Tribunal de Justiça.
UfEstado.csv	UfEstado.csv	IBGE	Possui dados do nome do estado e unidade federativa.
AnoEleicao.csv	AnoEleicao.csv	Autoria Própria	Possui dados de ano e se foi um ano que teve eleições para governador e presidente.

Fonte: Elaborado pelos autores.

5.2.2 Banco de Dados

Após a definição dos arquivos fontes, foi realizado a criação dos bancos de dados SQL Server para *staging* e DW, respectivamente nomeados como STG e DWCNJ.

O banco de dados STG foi modelado de acordo com os scripts presentes no apêndice Y e possui as seguintes tabelas:

- a) anoeleicao.
- b) documentacao.
- c) indicadores.
- d) ufestados.
- e) variáveis.

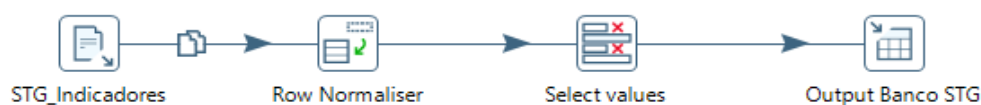
O banco de dados DWCNJ possui as tabelas do modelo dimensional de acordo com o apresentado na seção 4.2 e foi modelado de acordo com os scripts presentes no apêndice X.

5.2.3 Transformação Staging

As transformações *Staging* conforme explicado na subseção 2.2.7, são realizadas para tratar previamente os dados e armazená-los no mesmo banco de dados.

A tabela *stg_indicadores* foi criada para armazenar as informações presentes no arquivo Indicadores.csv disponibilizado pelo CNJ. Conforme a figura 14, antes de enviar os dados para a tabela *stg_indicadores*, foi necessário submetê-los a um *step* chamada *Row Normaliser*, que transforma colunas em linhas, dada a razão de que o número de colunas do arquivo fonte, ultrapassava o limite de 1024 colunas do SQL Server, em seguida, os dados são enviados para o Banco de Dados STG.

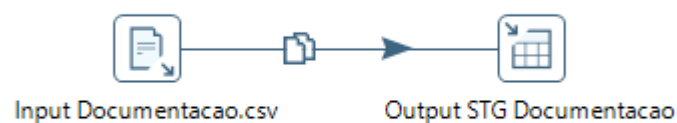
Figura 14 – Transformação Staging Indicadores



Fonte: Elaborado pelos autores.

As demais transformações *staging*, conforme a figura 15, possuem apenas *input* dos arquivos fonte .csv e *output* para o banco de dados STG.

Figura 15 – Transformação Staging Documentacao



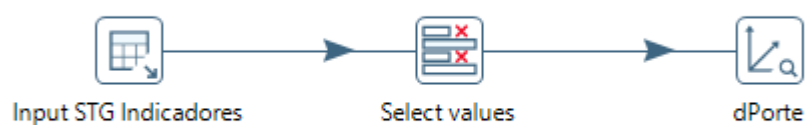
Fonte: Elaborado pelos autores.

Na próxima subseção, é apresentada informações sobre a transformação da dimensão porte.

5.2.4 Transformação dPorte

Conforme a figura 16, para a dimensão de porte do tribunal é realizado uma consulta distinta dos campos *idporte* e *porte* da tabela de Indicadores do banco STG, o código está ilustrado na figura 17, os campos são renomeados para respectivamente *cdPorte* e *nmPorte* e em seguida realizada a carga dos dados para a tabela dPorte do banco DWCNJ, onde serão adicionados os campos de controle *skPorte*, *date_from*, *date_to* e *version*.

Figura 16 – Transformação dPorte



Fonte: Elaborado pelos autores.

Figura 17 – Select Input dPorte

```

1  SELECT DISTINCT
2      idporte
3      ,   porte
4  FROM indicadores

```

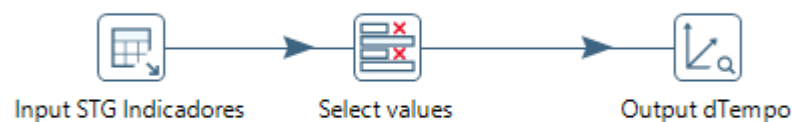
Fonte: Elaborado pelos autores.

Na próxima subseção, é apresentado informações sobre a transformação da dimensão tempo.

5.2.5 Transformação dTempo

Conforme a figura 18, para a dimensão de tempo é realizado uma consulta distinta dos campos *Ano* e *flEleicao*, respectivamente das tabelas de Indicadores e anoeleicao do banco STG, o código está ilustrado na figura 19, são renomeados para respectivamente *nuAno* e *flEleicao* e em seguida realizada a carga dos dados para a tabela dTempo do banco DWCNJ, onde serão adicionados os campos de controle *skTempo*, *date_from*, *date_to* e *version*.

Figura 18 – Transformação dTempo



Fonte: Elaborado pelos autores.

Figura 19 – Select Input dTempo

```

1 SELECT DISTINCT
2     i.ano
3     , CASE WHEN ae.ano IS NULL THEN 'N' ELSE 'S' END AS flEleicao
4 FROM indicadores i
5 LEFT JOIN anoeleicao ae ON ae.ano = i.ano
6

```

Fonte: Elaborado pelos autores.

Na próxima subseção, é apresentada informações sobre a transformação da dimensão tribunal.

5.2.6 Transformação dTribunal

Conforme a figura 22, para a dimensão de tribunal é realizado uma consulta dos campos *justiça*, *sigla*, *dsc_tribunal*, *uf_Sede* e *uf_abrangida* da tabela de Indicadores e os campos *unidade* e *UF* da tabela ufestados do banco STG respectivamente ilustrado pela figura 20 e figura 21, o campo *dsc_tribunal* é dividido em dois novos campos: *Tipo Tribunal* e *Regiao*.

Figura 20 – Select Input dTribunal

```

1 SELECT DISTINCT
2     justica
3     , sigla
4     , dsc_tribunal
5     , uf_sede
6     , uf_abrangida
7 FROM indicadores
8

```

Fonte: Elaborado pelos autores.

Figura 21 – Select Input dTribunal 2

```

1 SELECT
2     UF
3     , Unidade
4 FROM ufestados
5

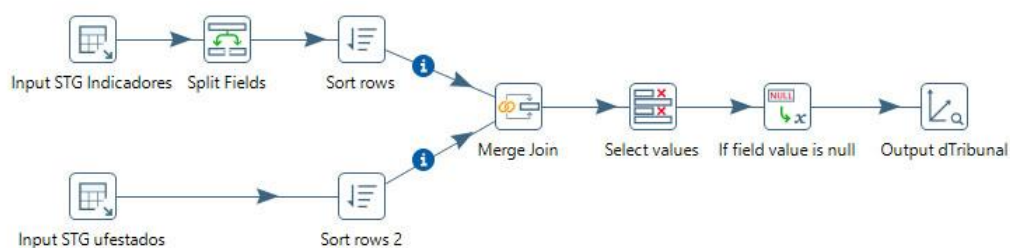
```

Fonte: Elaborado pelos autores.

É realizado uma junção entre as duas tabelas pelos campos *UF* e *uf_sede*, após isso são selecionados os campos *justica*, *sigla*, *Tipo Tribunal*, *uf_sede*, *Unidade* e *uf_abrangida* e renomeados respectivamente para *nmJustica*, *cdTribunal*, *deTipo*, *nmUF*, *nmEstado*, *deAbrangencia*.

Valores nulos na tabela são tratados e em seguida realizada a carga dos dados para a tabela *dTribunal* do banco *DWCNJ*, onde serão adicionados os campos de controle *skTribunal*, *date_from*, *date_to* e *version*.

Figura 22 – Transformação dTribunal



Fonte: Elaborado pelos autores.

Na próxima subseção, é apresentado informações sobre a transformação da dimensão indicador.

5.2.7 Transformação dIndicador

Conforme a figura 25, para a dimensão de indicador é realizado uma consulta dos campos *categoria*, *Sub Categoria* e *Chave* da tabela de documentacao e os campos *sigla* e

dsc_sigla da tabela de variáveis do banco STG, respectivamente ilustrado pela figura 23 e figura 24, é realizado uma junção entre as duas tabelas pelo campos *sigla* e *Chave*.

Figura 23 – Select Input dIndicador

```

1  SELECT distinct
2      UPPER(Categoria) AS Categoria
3      , UPPER("Sub Categoria") AS "Sub Categoria"
4      , Chave
5  FROM documentacao
6

```

Fonte: Elaborado pelos autores.

Figura 24 – Select Input dIndicador 2

```

1  SELECT
2      sigla
3      , dsc_sigla
4  FROM variaveis
5

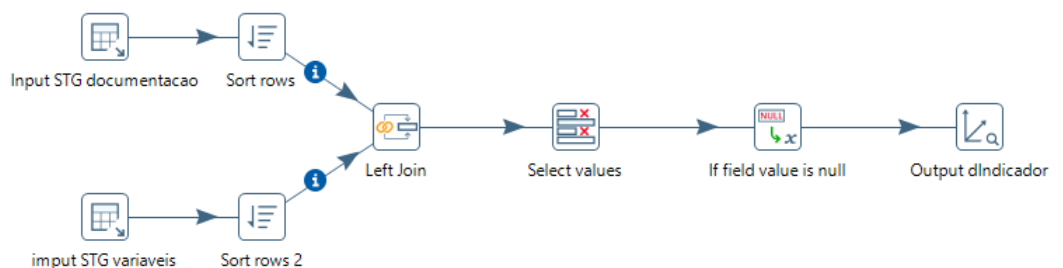
```

Fonte: Elaborado pelos autores.

Após isso são selecionados os campos *sigla*, *dsc_sigla*, *Categoria* e *Sub Categoria* e renomeados respectivamente para *cdIndicador*, *nmIndicador*, *nmCategoria* e *nmSubcategoria*.

Valores nulos na tabela são tratados e em seguida realizada a carga dos dados para a tabela dIndicador do banco DWCNJ, onde serão adicionados os campos de controle *skIndicador*, *date_from*, *date_to* e *version*.

Figura 25 – Transformação dIndicador



Fonte: Elaborado pelos autores.

Na próxima subseção, é apresentado informações sobre a transformação da tabela fato indicador.

5.2.8 Transformação fIndicador

Conforme a figura 27, para a fato de indicadores é realizado uma consulta dos campos *ano*, *sigla*, *indicadores*, *idporte* e *valor* da tabela de indicadores do banco STG, o código está ilustrado na figura 26, e renomeados para respectivamente os campos *cdAno*, *cdTribunal*, *cdIndicador*, *cdPorte* e *vlIndicador*.

Figura 26 – Select Input fIndicador

```

1  SELECT
2      ano
3      , sigla
4      , Indicadores
5      , idporte
6      , Valor
7  FROM indicadores
8

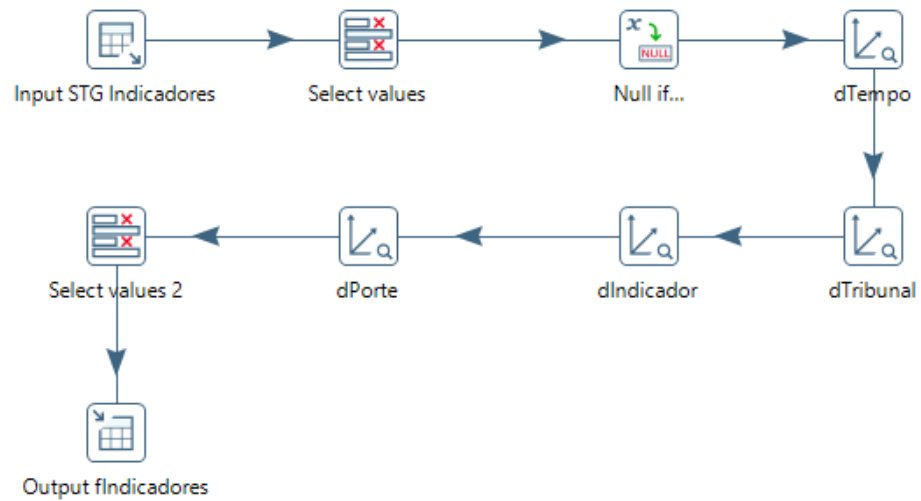
```

Fonte: Elaborado pelos autores.

Valores do tipo texto no campo *vlIndicador* são substituídos por nulo, após isso, utiliza-se o *steps* chamado *Dimension Lookup / Update*, que permite por meio da chave natural, trazer o último registro da *surrogate key* das dimensões dTempo, dTribunal, dIndicador e dPorte,

São removidos os campos *cdAno*, *cdTribunal*, *cdIndicador* e *cdPorte* e por último os dados são enviados para a tabela fIndicadores do banco DWCNJ

Figura 27 – Transformação fIndicador



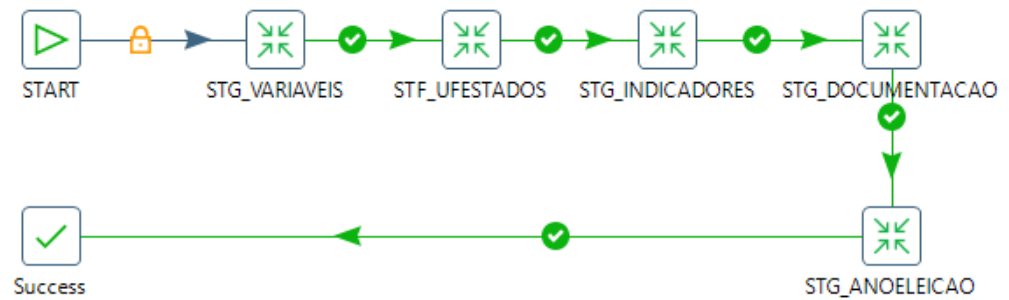
Fonte: Elaborado pelos autores.

Na próxima subseção, é apresentada informações sobre os *Jobs* utilizados no desenvolvimento.

5.2.9 Job Staging

O *Job Staging* irá transferir os dados .csv para tabelas no banco de dados STG, conforme a figura 28, em sequência, o *job* executará as transformações das tabelas variaveis, ufestados, indicadores, documentacao e anoeleicao.

Figura 28 – Job Staging



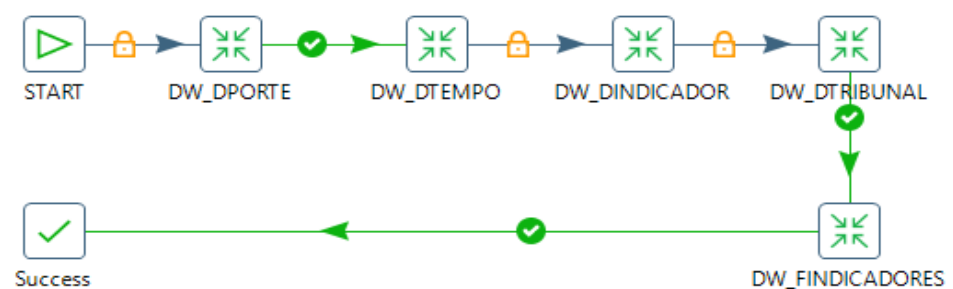
Fonte: Elaborado pelos autores.

Na próxima subseção, é apresentado o processo de *Job DW*, explicando o seu funcionamento e utilização.

5.2.10 Job DW

O *Job DW* irá transferir das tabelas variaveis, ufestados, indicadores, documentacao e anoeleicao do banco STG para tabelas no banco de dados DW, conforme a figura 29, em sequência, o *job* executará as transformações das tabelas dimensão dPorte, dTempo, dIndicador, dTribunal e fIndicadores.

Figura 29 – Job DW



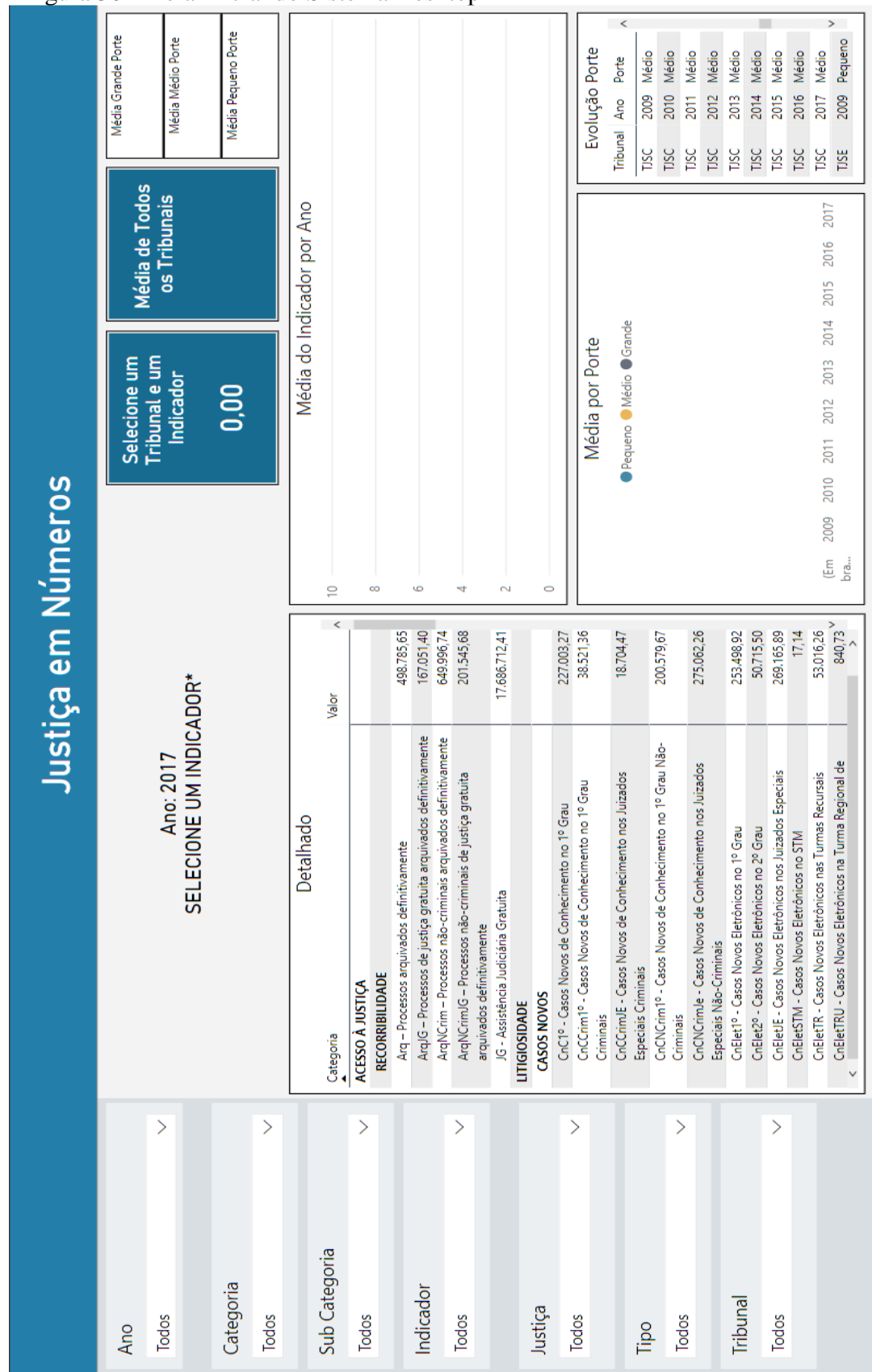
Fonte: Elaborado pelos autores.

Na próxima subseção, é apresentado o processo do Power BI, explicando o seu funcionamento e utilização.

5.2.11 Protótipo

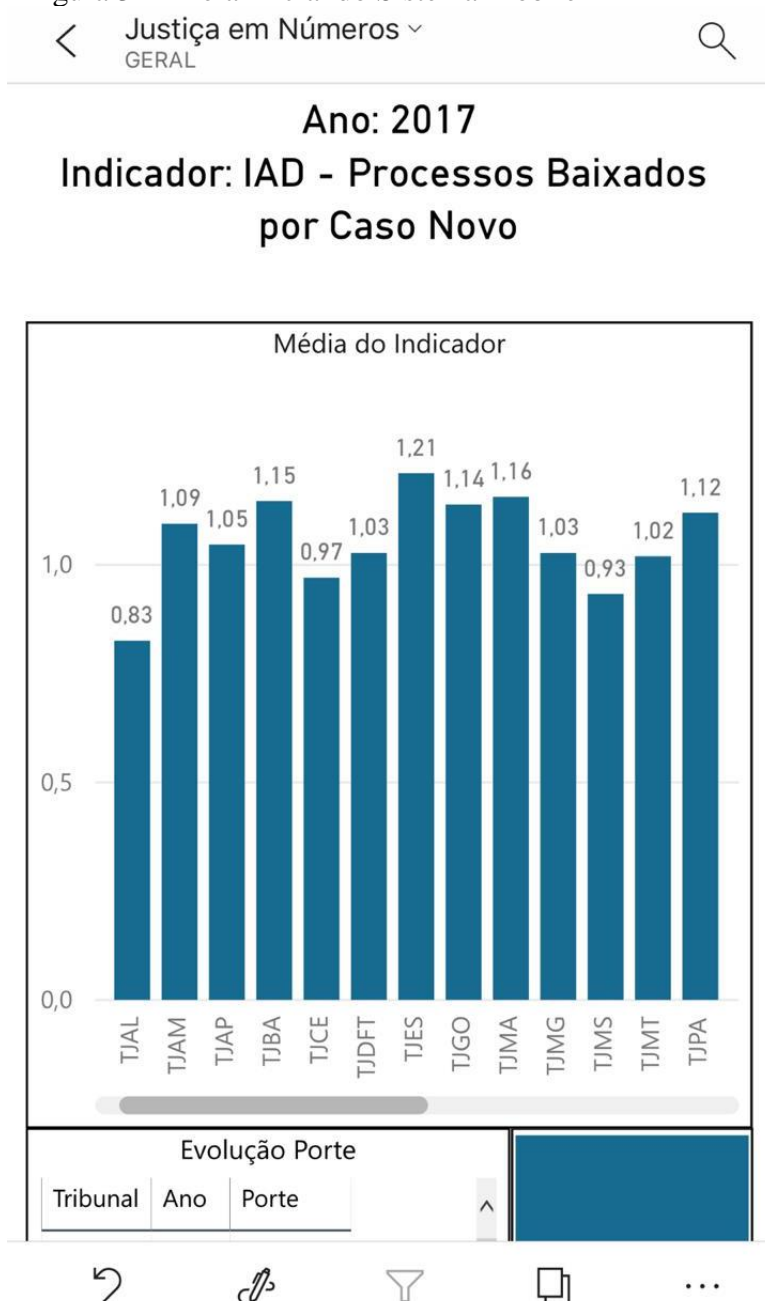
Conforme dito na subseção 5.1.1 a ferramenta escolhida para visualização de dados foi o Microsoft Power BI, a figura 30 ilustra a tela inicial do sistema na versão *desktop* e a figura 31 ilustra a tela de versão *mobile*.

Figura 30 – Tela Inicial do Sistema Desktop



Fonte: Elaborado pelos autores.

Figura 31 – Tela Inicial do Sistema Mobile



Fonte: Elaborado pelos autores.

Por padrão os dados exibidos são sempre do último ano disponível ou selecionado. Todos os elementos gráficos interagem entre si como filtros, ao clicar em um dado dentro de um elemento gráfico é possível filtrar um outro elemento gráfico.

Para melhor entendimento do usuário final, foi realizada a alteração do nome de alguns campos do modelo dentro do Power BI conforme o quadro 8, demais campos não citados no quadro e presentes no modelo dimensional da figura 11 foram ocultados.

Quadro 8 – Campos Power BI

Nome Origem	Novo Campo
nmCategoria	Categoria
cdIndicador	Código Indicador
nmIndicador	Indicador
nmSubcategoria	Sub Categoria
nmPorte	Porte
nuAno	Ano
flEleicao	Teve Eleição?
deAbragendo	Abrangência
nmEstado	Estado
nmJustica	Justiça
deTipo	Tipo
cdTribunal	Tribunal
nmUF	UF

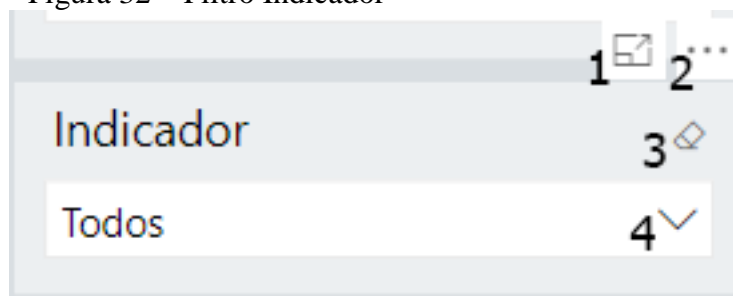
Fonte: Elaborado pelos autores.

As medidas *Data Analysis Expressions* (DAX) criadas dentro do Power BI estão no apêndice D.

5.2.11.1 Filtros

Os elementos gráficos de filtro, localizados no lado esquerdo da figura 30, possuem a funcionalidade de segmentar os dados.

Figura 32 – Filtro Indicador



Fonte: Elaborado pelos autores.

Ilustrado pela figura 32, o botão representado pela legenda 1 permite ativar o modo de foco, o qual expande o elemento gráfico em tela cheia, o botão representado pela legenda 2 permite utilizar funções de pesquisa, exportar dados, mostrar dados, destaque, classificar em ordem decrescente e classificar em ordem crescente, o botão representado pela legenda 3 permite limpar seleções aplicadas e o botão representado pela legenda 4 permite visualizar e selecionar dados para filtrar.

5.2.11.2 Cards

Os *cards*, localizados na parte superior da figura 30 trazem de forma direta valores do indicador previamente selecionado, há duas formas de visualizar os dados, filtrando apenas um indicador ou filtrando um indicador e um tribunal. Na figura 33 foi filtrado apenas o indicador: Arq – Processos arquivados definitivamente.

Figura 33 – Cards com Indicador Filtrado



Fonte: Elaborado pelos autores.

Ainda de acordo com a figura 33, no *card* representado pela legenda 1 solicita que seja filtrado um tribunal e um indicador, o *card* representado pela legenda 2 tem como informação a média do valor do indicador selecionado de todos os tribunais e o *card* representado pela legenda 3 separa a informação do *card* de legenda 2 por porte do tribunal.

Figura 34 – Cards com Indicador e Tribunal Filtrados



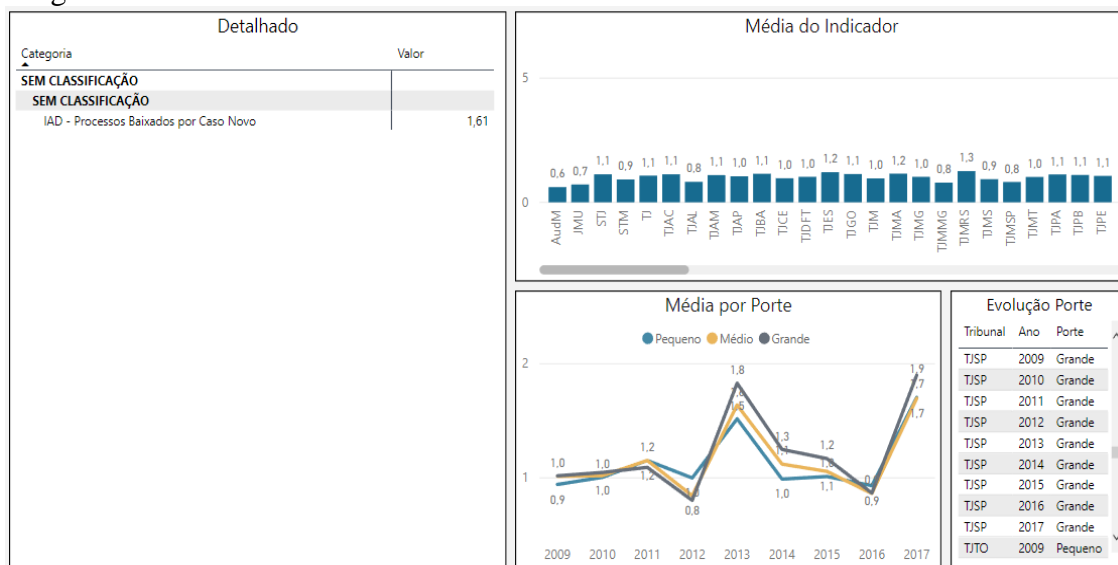
Fonte: Elaborado pelos autores.

Ao filtrar um indicador e um tribunal, conforme a figura 34 é possível ter a outra visão dos dados. No *card* representado pela legenda 1 é apresentado o valor do indicador para o tribunal filtrado, o *card* representado pela legenda 2 tem como informação a média do valor do indicador selecionado de todos os tribunais exceto o tribunal filtrado e o *card* representado pela legenda 3 separa a informação do *card* de legenda 2 por porte do tribunal.

5.2.11.3 Gráficos Principais

Os elementos gráficos localizados na parte central da figura 23 e conforme a figura 27 possuem os dados detalhados dos indicadores.

Figura 35 – Gráficos



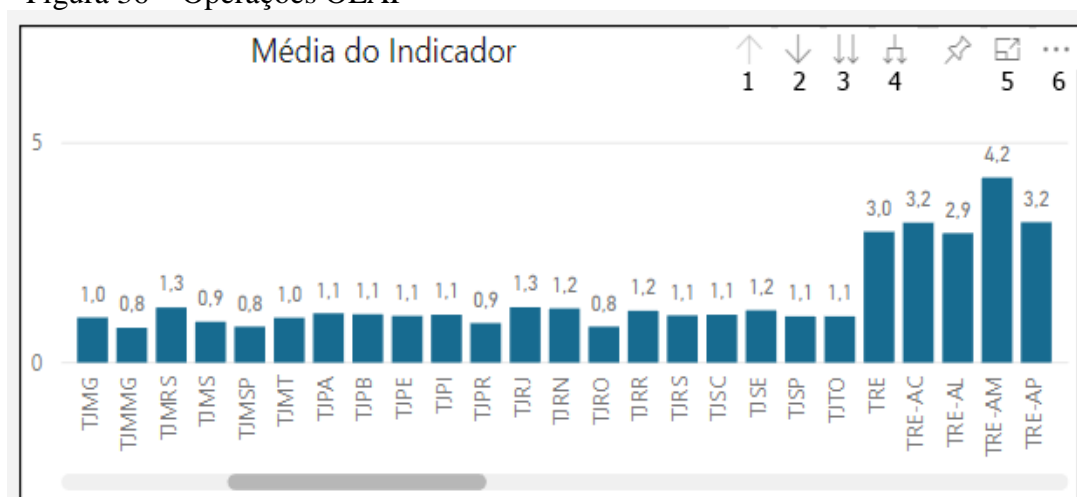
Fonte: Elaborado pelos autores.

De acordo com a figura 35, os elementos gráficos estão divididos em:

- Detalhado: possui dados da média do valor do indicador pela seguinte hierarquia de dados: Categoria, Subcategoria, Indicador e Tribunal.
- Média do Indicador: possui dados da média do valor do indicador pela seguinte hierarquia de dados: Tribunal e Ano.
- Média por Porte: possui dados da média do indicador por porte, caso haja um tribunal selecionado, os dados do mesmo não são calculados para média do porte correspondente.
- Evolução Porte: Situação anual da classificação do porte do tribunal.

Os elementos gráficos Detalhado e Média do Indicador possui as opções de realizar operações OLAP de acordo com a subseção 2.5.1.

Figura 36 – Operações OLAP



Fonte: Elaborado pelos autores.

Conforme a figura 36, o botão representado pela legenda 1 permite a operação *drill up*, o botão representado pela legenda 2 permite a operação *drill down*, o botão representado pela legenda 3 permite a operação *drill across*, o botão representado pela legenda 4 permite a operação *drill through*, o botão representado pela legenda 5 permite ativar o modo de foco, o qual expande o elemento gráfico em tela cheia e o botão representado pela legenda 6 permite utilizar funções de exportar dados, mostrar dados, destaque e opções de classificação.

As perguntas estratégicas definidas na subseção 4.1.1 são respondidas conforme o quadro 9 de acordo com os elementos gráficos da figura 30.

Quadro 9 – Perguntas Estratégicas x Elementos Gráficos

Pergunta estratégica	Elemento(s) gráfico(s)
Quais são os indicadores?	Detalhado.
Qual é o histórico do indicador selecionado?	Média do Indicador por Ano.
Qual o porte atual do tribunal e seu histórico?	Evolução Porte.
Qual é a média do indicador selecionado?	Cards, Média por Porte, Média do Indicador por Ano e detalhado.
Comparativo do tribunal selecionado com outros tribunais do mesmo porte.	Cards.

Qual a média de todos os outros tribunais desse indicador?	<i>Cards.</i>
Qual a média por porte do indicador selecionado?	<i>Cards.</i>
Qual a categorização e subcategorização de um determinado indicador?	Detalhado.
Quais são os tipos de tribunais disponíveis para análise?	Filtro Tipo.

Fonte: Elaborado pelos autores.

Na próxima seção, é apresentado a avaliação de entrevista com o *stakeholder*.

5.3 AVALIAÇÃO ENTREVISTA STAKEHOLDER

Tassiane Kolusso Palaoro é uma especialista da área de negócios da empresa Softplan, empresa que fornece sistemas que buscam automatizar os tribunais de justiça. Tassiane possui familiaridade com os dados e indicadores disponibilizados pelo CNJ, pois tem como atividade constante, o contato com os tribunais e escritórios de advocacia, com o objetivo de melhorar seus indicadores por meio de análise de dados.

Na compreensão da especialista, o relatório criado por meio de visualização gráfica neste trabalho é válido, por ser tratar de análises de dados que são constantemente solicitadas e analisadas pelos tribunais de justiça e advogados.

5.4 CONSIDERAÇÕES

Durante o desenvolvimento da solução foi utilizada a teoria descrita no capítulo 2 e foi possível identificar que com o uso de ferramentas como PDI e Power BI, assim como o conhecimento técnico auxiliou e facilitou a análise de dados de forma intuitiva e transparente.

Apesar do desafio de colocar em prática a teoria aprendida durante o desenvolvimento, os autores identificaram como maior dificuldade o entendimento dos indicadores e regras de negócio, pois eram mais de mil indicadores, e além disso, os dados disponibilizados não estavam em uma estrutura que permitisse o fácil entendimento e análise dos dados.

Uma outra grande dificuldade identificada foi a classificação dos indicadores. O CNJ disponibiliza uma documentação dos indicadores com suas classificações e subclassificações, no entanto, não estavam em um formato que permitia a combinação dessas documentações aos dados de indicadores disponibilizados. Foi necessário em um primeiro momento um trabalho manual dentro de preenchimento para que, em seguida, pudesse ser automatizado com a ajuda da ferramenta Excel.

É interessante notar que de 801 linhas do arquivo de origem, após o processo de normalização, se transformou em 1.102.176 linhas, com 1.221 indicadores e 97 tribunais distintos.

Apesar das dificuldades encontradas durante o desenvolvimento, a estrutura criada atende os requisitos definidos, os autores conseguiram entregar mais valor à solução com o trabalho de classificação dos indicadores, que permitiu uma melhor organização e interpretação dos indicadores.

6 CONCLUSÕES E TRABALHOS FUTUROS

Este capítulo irá apresentar as conclusões obtidas com o desenvolvimento do projeto, assim como sugestões de trabalhos futuros que poderão ser iniciados para complementar este trabalho.

6.1 CONCLUSÃO

Neste trabalho, foi desenvolvido uma solução de BI a partir dos dados abertos sobre os tribunais de justiça fornecidos pelo Conselho Nacional de Justiça e outras fontes. A ideia inicial desse projeto nasceu pelo motivo dos autores notarem que são informações relevantes para a população, porém a estrutura disponibilizada pelo órgão estão desnormalizados e com nomenclaturas complexas, sem padronização, dificultando o entendimento e o acesso às informações para o usuário final.

Criar visualizações e análises para um público alvo vasto é uma atividade complexa, pois é necessário generalizar, todavia é preciso conter informações precisas, claras e importantes, por esse motivo, foi criada a seção de perguntas estratégicas para que o projeto não fugisse do escopo. Essas perguntas foram utilizadas como base para os requisitos e regras, e como consequência, para a modelagem e para os relatórios.

Todo o desenvolvimento da solução foi feito utilizando metodologias e conceitos descritos no trabalho com e após a conclusão dessa etapa, os autores conseguiram realizar todos os passos planejados e de acordo com os requisitos.

O grande desafio do trabalho segundo os autores, foi o entendimento da regra de negócio, pois se trata de dados específicos voltado a justiça e saúde dos tribunais judiciários. Além de possuir mais de mil indicadores, não existe uma documentação clara e de fácil entendimento disponível no portal da fonte de dados, por esse motivo a tarefa de agrupar os indicadores em categorias e subcategorias para facilitar a organização e compreensão do usuário final foi demorada e complexa.

Com as regras e requisitos bem definidos foi criado um relatório utilizando a ferramenta Power BI da Microsoft para a visualização das informações, nela, os autores puderam validar as informações e dados dos valores dos indicadores, hierarquias, agrupamentos, filtros e publicar *online* de uma forma rápida e intuitiva.

Visto isso, é possível afirmar que este projeto cumpriu com a proposta definida e que a utilização de conceitos e ferramentas de BI funcionam de forma descrita em aplicações reais e fica evidente que um projeto de *Business Intelligence* é um trabalho de constante melhorias e evoluções.

6.2 TRABALHOS FUTUROS

A partir deste trabalho é possível desenvolver aperfeiçoamentos na proposta apresentada, criação de novas ideias, exploração de novas ferramentas e criação de novas análises, pois trata-se de dados atualizados anualmente pelo CNJ. A informação apresentada neste repositório possui muita relevância para o dia-a-dia dos cidadãos brasileiros por se tratar de dados e indicadores sobre a saúde dos tribunais, podendo ser utilizada e analisada de diversas formas.

A primeira sugestão de trabalhos futuros é a automatização da coleta de dados fornecidos pelo Conselho Nacional de Justiça em formatos *.csv*, *web*, *.xlsx*. Neste trabalho a extração dos dados no *site* foi feita de forma manual, seja por *download* ou via tabelas *html*. A partir do momento em que a etapa de automatização for concluída, é possível agendar atualizações rotineiras e ETLs de forma contínua para que, independente do ano corrente, os dados estejam atualizados.

Análises de dados mais aprofundadas em relação aos indicadores também são um interessante trabalho futuro. Por se tratar de mais de mil indicadores, os tipos e opções de análises matemáticas e estatísticas possíveis são vastas, desde um simples cálculo até análises preditivas e prescritivas.

Uma nova ampliação de escopo seria a criação de novas abas e relatórios no Power BI ou uma outra ferramenta adequada, baseada em ideias e sugestões de analista de negócios,

jornalistas, advogados ou de qualquer pessoa que veja a necessidade de mais visualizações. Além disso, é possível acrescentar no modelo de dados mais informações de outras fontes com o objetivo de enriquecer as análises necessárias, alguns exemplos de fontes sugeridas pelos autores são o IBGE, Justica.gov, Tribunais Regionais, entre outros.

Concluindo, a última sugestão de um trabalho futuro seria a coleta de dados via *crawlers* dos dados públicos de processos judiciais abertos, disponíveis para a população, para que com isso se possa comparar a qualidade dos dados que são fornecidos pelo Conselho Nacional de Justiça.

REFERÊNCIAS

ADAMSON, Christopher. **Mastering Data Warehouse Aggregates: Solutions for Star Schema Performance**. 2006.

ADAMSON, Christopher. **Star Schema The Complete Reference**. 2010.

ANGELONI, Maria Terezinha. **Elementos Intervenientes na Tomada de Decisão**. Ci. Inf., Brasília, v. 32, n. 1, jan./abr. 2003.

BARBIERI, Carlos. **BI2 – Business Intelligence: modelagem e qualidade**. Rio de Janeiro. Elsevier, 2011.

BRASIL. **Lei Complementar nº 131, de 27 de maio de 2009**. Disponível em: <http://www.planalto.gov.br/CCivil_03/leis/LCP/Lcp131.htm> Acesso em: 30 mar. 2018.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011**. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/12527.htm> Acesso em: 30 mar. 2018.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm> Acesso em: 20 maio 2018.

CECI, Flávio. **Business Intelligence**. Palhoça : UnisulVirtual, 2012.

CONSELHO DA JUSTIÇA FEDERAL, Conselho da Justiça Federal. **Primeira instância, segunda instância... Quem é quem na Justiça brasileira?** 2010. Disponível em: <<http://cnj.jus.br/noticias/cnj/59220-primeira-instancia-segunda-instancia-quem-e-quem-na-justica-brasileira>>. Acesso em: 19 maio. 2018.

CONSELHO DA JUSTIÇA FEDERAL, Conselho da Justiça Federal. **Especialistas ressaltam importância do levantamento de dados sobre o Judiciário brasileiro**, 2015. Disponível em: <<http://www.cjf.jus.br/cjf/noticias/2015/setembro/especialistas-ressaltam-importancia-do-levantamento-de-dados-sobre-o-judiciario-brasileiro>>. Acesso em: 30 mar. 2018.

CONSELHO NACIONAL DE JUSTIÇA, Conselho Nacional de Justiça. **Justiça em números 2017**. Brasília, 2017. Disponível em: <<http://www.cnj.jus.br/files/conteudo/arquivo/2017/12/b60a659e5d5cb79337945c1dd137496c.pdf>>. Acesso em: 18 mar. 2018.

CONSELHO NACIONAL DE JUSTIÇA. **Quem somos**, [2009?]. Disponível em: <<http://www.cnj.jus.br/sobre-o-cnj/quem-somos-visitas-e-contatos>>. Acesso em: 19 maio. 2018.

FAYYAD, Usama; GREGORY, Piatetsky-Shapiro; SMYTH, Padhraic. **From Data Mining to Knowledge Discovery in Databases**. AI Magazine Volume 17 Number 3 1996. Disponível em: <<https://www.aaai.org/ojs/index.php/aimagazine/article/view/1230/1131>> Acesso em: 14 maio. 2018.

FOLHA DE SÃO PAULO. Para entender direito: **A história do CNJ**, 2012. Disponível em: <<http://direito.folha.uol.com.br/blog/a-histria-do-cnj>>. Acesso em: 19 maio. 2018.

GARTNER. **Magic Quadrant for Analytics and Business Intelligence Plataforms**, 2018. Disponível em: <<https://www.gartner.com/doc/3861464/magic-quadrant-analytics-business-intelligence>>. Acesso em: 16 outubro 2018.

GOVERNO DIGITAL. **Dados Abertos Governamentais**, 2016. Disponível em: <<https://www.governodigital.gov.br/eixos-de-atuacao/cidadao/dados-abertos/dados-abertos-governamentais>> Acesso em: 30 mar. 2018.

HARRISON, T. M. et al. **Open government and e-government: Democratic challenges from a public value perspective**, Information Polity, v. 17, p. 83-97, 2012.

IMHOFF, Claudia; GALEMMO, Nicholas; GEIGER, G, Jonathan. **Matering Data Warehouse Design: Relational and Dimensional Techniques**. Wiley Publishing, Canada 2003.

INMON, W. H. **Building the Data Warehouse, Fourth Edition**. Wiley Publishing Inc, Indianapolis 2005.

KIMBALL, Ralph. **Fact Tables**. Disponível em <<https://www.kimballgroup.com/2008/11/fact-tables/>>. Acesso em: 19 maio 2018.

KIMBALL, Ralph et al. **The Data Warehouse Lifecycle Toolkit**. Wiley Publishing Inc, Indianapolis, 1998.

KIMBALL, Ralph; CASERTA, Joe. **The Data Warehouse ETL Toolkit**. Wiley Publishing Inc, Indianapolis, 2004.

KIMBALL, Ralph; Ross, Margy. **The Data Warehouse Toolkit Second Edition The Complete Guide to Dimensional Modeling**. Wiley Publishing Inc, Indianapolis, 2002.

KLEIN, Rodrigo Hickmann.; LUCIANO, Edimara Mezzomo; MACADAR, Marie Anne. Grau de Transparência de Dados Abertos Governamentais do Site dados.rs.gov.br. **Revista Economia & Gestão**. Belo Horizonte, 2015.

LUHN, H. P. **A Business Intelligence System**, 1958. Disponível em: <<http://altaplana.com/ibmrd0204H.pdf>> Acesso em: 05 de maio. 2018.

MACHADO, Felipe Nery Rodrigues. **Tecnologia e Projeto de Data Warehouse: uma visão multidimensional**. 6 ed. São Paulo, Érica, 2013.

MICROSOFT. **Noções básicas sobre cubos OLAP**, 2016a. Disponível em: <[https://msdn.microsoft.com/pt-br/library/hh916543\(v=sc.12\).aspx](https://msdn.microsoft.com/pt-br/library/hh916543(v=sc.12).aspx)> Acesso em: 19 de maio. 2018.

MICROSOFT. **Sobre cubos OLAP**, 2016b. Disponível em: <[https://docs.microsoft.com/pt-br/previous-versions/system-center/system-center-2012-R2/hh916536\(v=sc.12\)](https://docs.microsoft.com/pt-br/previous-versions/system-center/system-center-2012-R2/hh916536(v=sc.12))> Acesso em: 22 de maio. 2018.

OPEN GOVERNMENT PARTNERSHIP, Open Government Partnership. **Open Government Declaration**, 2011. Disponível em: <<https://www.opengovpartnership.org/open-government-declaration>>. Acesso em: 30 mar. 2018.

PITON, Rafael. **Data Warehouse Passo a Passo: O Guia Prático de Como Construir um Data Warehouse do Zero**. Porto Alegre, 2018.

POWER, Daniel J. **A Brief History of Decision Support Systems**. 2007. Disponível em: <<http://dssresources.com/history/dsshistory.html>> Acesso em: 05 de maio. 2018.

PRIMAK, F.V. **Decisões com BI (Business Intelligence)**. 2001.

PUJARI, Arun K. **Data Mining Techniques**. Rio de Janeiro: Ciência Moderna, 2008.

ROVER, Aires José. **Engenharia e gestão do judiciário brasileiro: Estudos sobre E-Justiça**. Erechim, 2016.

SANDERS, G. Lawrance; SHIN, Seungkyoon. **Denormalization Effects on Performance of RDBMS**. Hawaii International Conference on System Sciences, 2001. Disponível em: <<https://pdfs.semanticscholar.org/2c79/069c01ba8d598f32e61fe367ef6d261a0cb4.pdf>> Acesso em: 18 mar. 2018.

SELL, Denilson. **Uma arquitetura para business intelligence baseada em tecnologias semânticas para suporte a aplicações analíticas**. Florianópolis, 2006.

SILVA, E. L.; MENEZES, E. M. **Metodologia da Pesquisa e Elaboração de Dissertação**. 4ª. ed. Florianópolis: UFSC, 2005.

SOMMERVILLE, Ian. **Engenharia de Software**. 9. ed. São Paulo: Pearson Prentice Hall, 2011.

SUPREMO TRIBUNAL FEDERAL, Supremo Tribunal Federal. **Notícias STF**, 2017. Disponível em: <<http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=346188>>. Acesso em: 30 mar. 2018.

TURBAN, Efraim; SHARDA, Ramesh; ARONSON, Jey E.; KING, David. **Business Intelligence: Um Enfoque Gerencial para Inteligência do Negócio**. 2009. Tradução Fabiano Bruno Gonçalves.

UBALDI, Barbara. **Open Government Data: Toward Empirical Analysis of Open Government Data Initiatives**. OECD Working Papers on Public Governance. Paris, 2013. Disponível em <<https://www.oecd-ilibrary.org/docserver/5k46bj4f03s7-en.pdf?expires=1522683097&id=id&accname=guest&checksum=2FE39FEFF35DDAA34E015D4BF2AD6EB1>>. Acesso em: 01 abr. 2018.

VAZQUEZ, Carlos Eduardo; SIMÕES, Guilherme Siqueira, **Engenharia de requisitos: Software orientado ao negócio**, Rio de Janeiro: Brasport, 2016.

VELTEM, Fernando Lugão. **Snowflake Schema**, 2012. Disponível em: <<https://social.technet.microsoft.com/wiki/contents/articles/8817.snowflake-schema-pt-br.aspx>> Acesso em: 13 maio. 2018

ZAKER, Morteza et al. Hierarchical Denormalizing: A Possibility to Optimize the Data Warehouse Design. **International Journal of Computers**, 2009. Disponível em <https://www.researchgate.net/publication/242072514_Hierarchical_Denormalizing_A_Possibility_to_Optimize_the_Data_Warehouse_Design> Acesso em: 18 mar. 2018.

APÊNDICES

APÊNDICE B – DDL STG

```
CREATE TABLE variaveis (  
    sigla VARCHAR(30),  
    dsc_sigla VARCHAR(1000)  
);  
  
CREATE TABLE ufestados (  
    UF VARCHAR(100),  
    Unidade VARCHAR(100),  
    Codigo VARCHAR(100)  
);  
  
CREATE TABLE indicadores (  
    ano VARCHAR(5),  
    justica VARCHAR(1000),  
    sigla VARCHAR(10),  
    dsc_tribunal VARCHAR(1000),  
    uf_sede VARCHAR(7),  
    uf_abrangida VARCHAR(1000),  
    seq_orgao VARCHAR(1000),  
    Indicadores VARCHAR(1000),  
    Valor VARCHAR(1000),  
    estrutura VARCHAR(100),  
    idporte VARCHAR(100),  
    porte VARCHAR(100)  
);  
  
CREATE TABLE documentacao (  
    Tribunal VARCHAR(255),  
    Categoria VARCHAR(1000),  
    Sub Categoria VARCHAR(1000),  
    Indicador VARCHAR(1000),  
    Chave VARCHAR(255)  
);  
  
CREATE TABLE anoEleicao (  
    ano BIGINT  
);
```


APÊNDICE C – DDL DWCNJ

```

CREATE TABLE dTribunal (
    skTribunal BIGINT NOT NULL,
    version INTEGER,
    date_from DATE,
    date_to DATE,
    cdTribunal VARCHAR(10),
    nmJustica VARCHAR(1000),
    deTipo VARCHAR(100),
    nmUF VARCHAR(7),
    nmEstado VARCHAR(100),
    deAbrangencia VARCHAR(1000),
    CONSTRAINT PK__dTribuna__F9A51E0C413D6A4C PRIMARY KEY
(skTribunal)
);

CREATE INDEX idx_dTribunal_lookup
ON dTribunal
( cdTribunal ASC );

CREATE INDEX idx_dTribunal_tk
ON dTribunal
( skTribunal ASC );

CREATE TABLE dTempo (
    skAno BIGINT NOT NULL,
    version INTEGER,
    date_from DATE,
    date_to DATE,
    nuAno VARCHAR(5),
    flEleicao VARCHAR(1),
    CONSTRAINT PK__dTempo__75C79E91DDF65CDA PRIMARY KEY (skAno)
);

CREATE INDEX idx_dTempo_lookup
ON dTempo
( nuAno ASC );

CREATE INDEX idx_dTempo_tk
ON dTempo
( skAno ASC );

CREATE TABLE dPorte (
    skPorte BIGINT NOT NULL,
    version INTEGER,
    date_from DATE,
    date_to DATE,
    cdPorte VARCHAR(100),
    nmPorte VARCHAR(100),
    CONSTRAINT PK__dPorte__956902D3D891D49B PRIMARY KEY
(skPorte)
);

CREATE INDEX idx_dPorte_lookup
ON dPorte
( cdPorte ASC );

```

```
CREATE INDEX idx_dPorte_tk
ON dPorte
( skPorte ASC );
```

```
CREATE TABLE dIndicador (
    skIndicador BIGINT NOT NULL,
    version INTEGER,
    date_from DATE,
    date_to DATE,
    cdIndicador VARCHAR(30),
    nmIndicador VARCHAR(1000),
    nmCategoria VARCHAR(1000),
    nmSubcategoria VARCHAR(1000),
    CONSTRAINT PK__dIndicad__52EAA386DC4740CE PRIMARY KEY
(skIndicador)
);
```

```
CREATE INDEX idx_dIndicador_lookup
ON dIndicador
( cdIndicador ASC );
```

```
CREATE INDEX idx_dIndicador_tk
ON dIndicador
( skIndicador ASC );
```

```
CREATE TABLE fIndicadores (
    skTribunal BIGINT NOT NULL,
    skAno BIGINT NOT NULL,
    skPorte BIGINT NOT NULL,
    skIndicador BIGINT NOT NULL,
    vlIndicador VARCHAR(100),
    CONSTRAINT fIndicadores_pk PRIMARY KEY (skTribunal, skAno,
skPorte, skIndicador)
);
```

```
ALTER TABLE fIndicadores ADD CONSTRAINT dTribunal_fIndicadores_fk
FOREIGN KEY (skTribunal)
REFERENCES dTribunal (skTribunal)
ON DELETE NO ACTION
ON UPDATE NO ACTION;
```

```
ALTER TABLE fIndicadores ADD CONSTRAINT dTempo_fIndicadores_fk
FOREIGN KEY (skAno)
REFERENCES dTempo (skAno)
ON DELETE NO ACTION
ON UPDATE NO ACTION;
```

```
ALTER TABLE fIndicadores ADD CONSTRAINT dPorte_fIndicadores_fk
FOREIGN KEY (skPorte)
REFERENCES dPorte (skPorte)
ON DELETE NO ACTION
ON UPDATE NO ACTION;
```

```
ALTER TABLE fIndicadores ADD CONSTRAINT dIndicador_fIndicadores_fk
FOREIGN KEY (skIndicador)
REFERENCES dIndicador (skIndicador)
ON DELETE NO ACTION
ON UPDATE NO ACTION;
```

APÊNDICE D – FORMULAS DAX

Indicador Selecionado =

```
IF ( HASEXIST ( dIndicador[Indicador] ); [Valor Indicador]; BLANK ( ) )
```

Média do indicador não selecionado de grande porte =

```
CALCULATE ( [Média do Não selecionado]; dPorte[Porte] = "Grande" )
```

Média do indicador não selecionado de médio porte =

```
CALCULATE ( [Média do Não selecionado]; dPorte[Porte] = "Médio" )
```

Média do indicador não selecionado de pequeno porte =

```
CALCULATE ( [Média do Não selecionado]; dPorte[Porte] = "Pequeno" )
```

Média do Não selecionado =

```
VAR valorSelecionado =
```

```
    IF (
        HASEXIST ( dTribunal[Tribunal] );
        SELECTEDVALUE ( dTribunal[Tribunal] )
    )
```

```
RETURN
```

```
    IF (
        HASEXIST ( dIndicador[Indicador] );
        CALCULATE ( [Valor Indicador]; dTribunal[Tribunal] <> valorSelecionado );
        " "
    )
```

Texto - Média dos Não Selecionados =

```
IF (
    HASEXIST ( dTribunal[Tribunal] );
    CONCATENATE (
        CONCATENATE (
            "Média de Todos os Tribunais (Exceto ";
            SELECTEDVALUE ( dTribunal[Tribunal] )
        );
    );
```

```

        ")"
    );
    "Média de Todos os Tribunais"
)
Texto - Selecione =
"Ano: " & MAX ( dTempo[Ano] ) & UNICHAR ( 10 ) & IF (
    HASONEFILTER ( dIndicador[Indicador] );
    "Indicador: " & MIN ( dIndicador[Indicador] );
    "SELECIONE UM INDICADOR*"
) & UNICHAR ( 10 ) & IF (
    HASONEFILTER ( dTribunal[Tribunal] );
    "Tribunal: " & MIN ( dTribunal[Tribunal] );
    ""
)
Texto - Selecione TJ =
IF (
    AND (
        HASONEFILTER ( dIndicador[Indicador] );
        HASONEFILTER ( dTribunal[Tribunal] )
    );
    "Valor do " & MIN ( dTribunal[Tribunal] );
    "Selecione um Tribunal e um Indicador"
)
Valor do Seleccionado =
VAR maxAno =
    CALCULATE (
        MAX ( dTempo[Ano] );
        FILTER (
            dTempo;
            MAX ( dTempo[Ano] )
        )
    )
RETURN

```

```

IF (
    AND (
        HASONEFILTER ( dIndicador[Indicador] );
        HASONEFILTER ( dTribunal[Tribunal] )
    );
    [Valor Indicador];
    ""
)
Valor Indicador =
VAR maxAno =
    CALCULATE (
        MAX ( dTempo[Ano] );
        FILTER (
            dTempo;
            MAX ( dTempo[Ano] )
        )
    )
RETURN
    CALCULATE (
        AVERAGE ( fIndicadores[vIndicador] );
        dTempo[Ano] = maxAno
    )

```