



CENTRO UNIVERSITÁRIO LUTERANO DE PALMAS

Recredenciado pela Portaria Ministerial nº 1.162, de 13/10/16, D.O.U. nº 198, de 14/10/2016
AELBRA EDUCAÇÃO SUPERIOR - GRADUAÇÃO E PÓS-GRADUAÇÃO S.A.

Thiago Aparecido da Silva

DESENVOLVIMENTO DE UM DATA MART E AUTOMATIZAÇÃO DO PROCESSO
ETL PARA O SISTEMA DE PROCEDIMENTOS EXTRAJUDICIAIS DO MPE/TO

Palmas – TO

2020

Thiago Aparecido da Silva

DESENVOLVIMENTO DE UM DATA MART E AUTOMATIZAÇÃO DO PROCESSO
ETL PARA O SISTEMA DE PROCEDIMENTOS EXTRAJUDICIAIS DO MPE/TO

Projeto de Pesquisa elaborado e apresentado como requisito parcial para aprovação na disciplina de Trabalho de Conclusão de Curso I (TCC I) do curso de bacharel em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientador: Prof. Esp. Fábio Castro Araújo.

Palmas – TO

2020

Thiago Aparecido da Silva

DESENVOLVIMENTO DE UM DATA MART E AUTOMATIZAÇÃO DO PROCESSO
ETL PARA O SISTEMA DE PROCEDIMENTOS EXTRAJUDICIAIS DO MPE/TO

Projeto de Pesquisa elaborado e apresentado como requisito parcial para aprovação na disciplina de Trabalho de Conclusão de Curso I (TCC I) do curso de bacharel em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientador: Prof. Esp. Fábio Castro Araújo.

Aprovado em: ____/____/____

BANCA EXAMINADORA

Prof. Esp. Fábio Castro Araújo

Orientador

Centro Universitário Luterano de Palmas – CEULP

Prof. M.e Jackson Gomes de Souza

Centro Universitário Luterano de Palmas – CEULP

Prof.a M.a Madianita Bogo Marioti

Centro Universitário Luterano de Palmas – CEULP

Palmas – TO

2020

AGRADECIMENTOS

Agradeço primeiramente a Deus por ter me iluminado e dado todas as condições para poder realizar mais um sonho na minha vida. Aos meus pais e irmão, que deram apoio e incentivos nas horas difíceis, sempre me ajudando a seguir em frente e a nunca desistir.

A minha família que sempre esteve ao meu lado contribuindo para que eu pudesse concluir essa etapa em minha vida.

Agradeço aos professores, que sempre estiveram presentes, ajudando em todos os momentos tanto para o crescimento pessoal como profissional.

A todos meus amigos e colegas aqueles que contribuíram de alguma maneira para que pudesse concluir esta caminhada. Agradeço também, a um grande amigo que fiz durante a graduação, Lucas Costa. Obrigado pela sua amizade e apoio.

RESUMO

SILVA, Thiago Aparecido. **Desenvolvimento de um Data Mart e automatização do processo ETL para o sistema de procedimentos extrajudiciais do MPE/TO**. 2020. 23f. Trabalho de Conclusão de Curso (Graduação) – Curso de Sistema de Informação, Centro Universitário Luterano de Palmas, Palmas/TO, 2020.

A todo instante, um enorme volume de informações sobre variados aspectos de uma organização é gerado e armazenado. Entretanto, esses dados estão espalhados por diversos sistemas, isto torna a centralização dos dados complexos. Para sanar essa necessidade surgiu o *Data Warehouse*, que contribui para manipulação de grande volume de dados e tem por objetivo dar suporte a esses processos. O objetivo deste trabalho é construir um *data mart*, que representa um subconjunto do *Data Warehouse*, em que serão armazenados dados extraídos do sistema de procedimentos extrajudiciais do Ministério Público do Tocantins. Além disso, será realizada a automatização do processo extração, transformação e carregamento (*ETL*) através da ferramenta *Kettle* para o *data mart* desenvolvido. Para tanto, este trabalho aborda, em um primeiro momento, conceitos relacionados à *Data Warehouse*, apresentando seus modelos de dados, arquitetura, implementação, características e aplicação do processo *ETL* (Extract, Transform and Load). Já no segundo momento, o trabalho apresenta a metodologia, sendo abordado os materiais utilizados e os procedimentos necessários para o desenvolvimento do projeto. Por último, os resultados obtidos foram a estrutura física do *data mart* implementada com as tabelas fato e as tabelas dimensões criadas para receber os dados, o processo *ETL* e a criação de *Jobs* para a automatização desses processos foram criadas.

Palavras-chave: *data warehouse*; *data mart*; *ETL*;

LISTA DE FIGURAS

Figura 1 - Arquitetura de um Data Warehouse.....	9
Figura 2 - Idealização do conceito de Integração.....	10
Figura 3 - Exemplo de um Esquema Estrela.....	12
Figura 4 - Exemplo de um Esquema Floco de Neve.....	13
Figura 5 - Implementação top down.....	15
Figura 6 - Implementação bottom up.....	15
Figura 7 - ETL.....	16
Figura 8 - Relacionamento dos Softwares.....	18
Figura 9 - Interface gráfica do Spoon.....	20
Figura 10 - Metodologia.....	21
Figura 11 - Modelo logico do data mart.....	23
Figura 12 - Modelo fisico do data mart.....	23
Figura 13 - Processo ETL da tabela dimensão promotor.....	27
Figura 14 - Processo ETL da tabela dimensão tempo.....	28
Figura 15 - Processo ETL da tabela fato procedimento.....	29
Figura 16 - Jobs.....	29

LISTA DE TABELAS

Tabela 1 - Data Mart x Data Warehouse.....	14
Tabela 2 - Fato Procedimento.....	24
Tabela 3 - Dimensão Promotor.....	24
Tabela 4 - Dimensão Região.....	25
Tabela 5 - Dimensão Área.....	25
Tabela 6 - Dimensão Tempo.....	26

LISTA DE ABREVIATURAS E SIGLAS

MPE/TO	Ministério Público do Estado do Tocantins
DW	<i>Data Warehouse</i>
DM	<i>Data Mart</i>
ETL	<i>Extraction Transformation and Load</i>
E-ext	<i>Procedimento Eletrônico Extrajudicial</i>
PDI	<i>Pentaho Data Integration</i>
SK	<i>Surrogate key</i>
PK	<i>Primary Key</i>

SUMÁRIO

1 INTRODUÇÃO	7
2 REFERENCIAL TEÓRICO	9
2.1 DATA WAREHOUSE	9
2.1.1 MODELO DOS DADOS	11
2.1.2 DATA MART	13
2.1.3 IMPLEMENTAÇÃO	14
2.2 PROCESSO ETL	16
3 METODOLOGIA	18
3.1 FERRAMENTAS	18
3.1.1 E-ext	18
3.1.2 Pentaho Data Integration	19
3.1.3 PostgreSQL	20
3.2 MÉTODOS	20
4 RESULTADOS E DISCUSSÃO	22
4.1 PROJETO DO MODELO LÓGICO	22
4.2 PROJETO DO MODELO FÍSICO	23
4.3 PROCESSO ETL	26
4.3.1 Processo <i>ETL</i> da tabela dimensão (DIM_PROMOTOR)	26
4.3.2 Processo <i>ETL</i> da tabela dimensão (DIM_TEMPO)	28
4.3.3 Carga da tabela fato (FATO_PROCEDIMENTO)	28
4.3.2 Execução das transformações do <i>Data Mart</i>	29
5 CONSIDERAÇÕES FINAIS	31
REFERÊNCIAS	32

1 INTRODUÇÃO

A todo momento, uma quantidade de dados expressivos são produzidos dentro de organizações, contendo informações sobre os mais variados aspectos. De acordo com estudo da Business Software Alliance (BSA), 2,5 quintilhões de bytes de dados são criados todos os dias. Em contrapartida, tais dados geralmente encontram-se espalhados em diferentes sistemas, hardwares e plataformas, não havendo entre eles nenhuma forma de integração e padronização.

De acordo com Kimbal (1996), *Data Warehouse* é um sistema utilizado para extrair, limpar, adaptar e entregar dados históricos, com diversos padrões, para um banco de dados dimensional com a finalidade de suportar análises baseadas em informações confiáveis e apoiar o processo de tomada de decisão. Em geral, um *Data Warehouse* compreende um grande repositório de dados que armazenam informações de diferentes sistemas de certa organização e com base em análise dos dados é possível ter um suporte à decisão.

Compreende-se também que o *Data Warehouse* é composto por *Data Marts*, ou seja, os espaços onde os dados são separados de forma mais específica e são acessados diretamente pelos usuários finais. Por exemplo, se uma organização possui um setor de recursos humanos e outro de vendas, cada setor irá possuir um *Data Mart* para o armazenamento de seus dados, estruturados para atender sua necessidade.

O desenvolvimento de um *Data Warehouse* é constituído de várias etapas, sendo o processo *ETL* (Extract, Transform and Load) uma das fases primordiais para manter a qualidade e integração dos dados. Este processo é composto por três etapas: extração, responsável em extrair os dados; transformação, encarregado para realizar a limpeza e as alterações necessárias nos dados; e carga, responsável por integrar os dados no ambiente de *Data Warehouse*.

Em se tratando de procedimentos extrajudiciais, segundo Ortega (2016), são processos que visam resolver conflitos sem acionar o poder judiciário, isto é, os conflitos podem ser resolvidos entre as partes com a mediação de um órgão responsável.

O Ministério Público do Estado do Tocantins (MPE/TO) dispõe de um sistema responsável pelo gerenciamento destes procedimentos, no qual é realizado o controle da instauração, tramitação e conclusão de processos extrajudiciais. Entretanto, os dados produzidos no sistema encontram-se armazenados em bases distintas, ocasionando, assim, um esforço para transformar os dados brutos e dispersos em uma única fonte de dados.

Desta forma, o problema está no processo de extração, transformação e carga dos dados, pois a forma em que está organizado não permite a obtenção de respostas para

perguntas mais complexas sobre esses dados. Quanto ao problema do processo de automatização *ETL*, faz-se necessário que o MPE disponha de tecnologias que agregam valor no direcionamento dos dados, sendo possível a unificação destas informações em um único ambiente de maneira padronizada.

Portanto, o objetivo deste trabalho é desenvolver um *data mart* e automatizar o processo de extração, transformação e carga para centralizar os dados referentes ao sistema de procedimentos extrajudiciais (*e-Ext*). Por meio da aplicação deste trabalho será possível utilizar tecnologias que ofereçam recursos de integração dos dados de uma maneira flexível e eficiente.

Este documento está organizado da seguinte forma: o segundo capítulo apresenta o referencial teórico necessário para o desenvolvimento do trabalho; o terceiro capítulo apresenta a metodologia e ferramentas adotadas; o quarto capítulo refere-se aos resultados e discussões obtidas durante o desenvolvimento e o quinto capítulo versa sobre as considerações finais e trabalhos futuros.

2 REFERENCIAL TEÓRICO

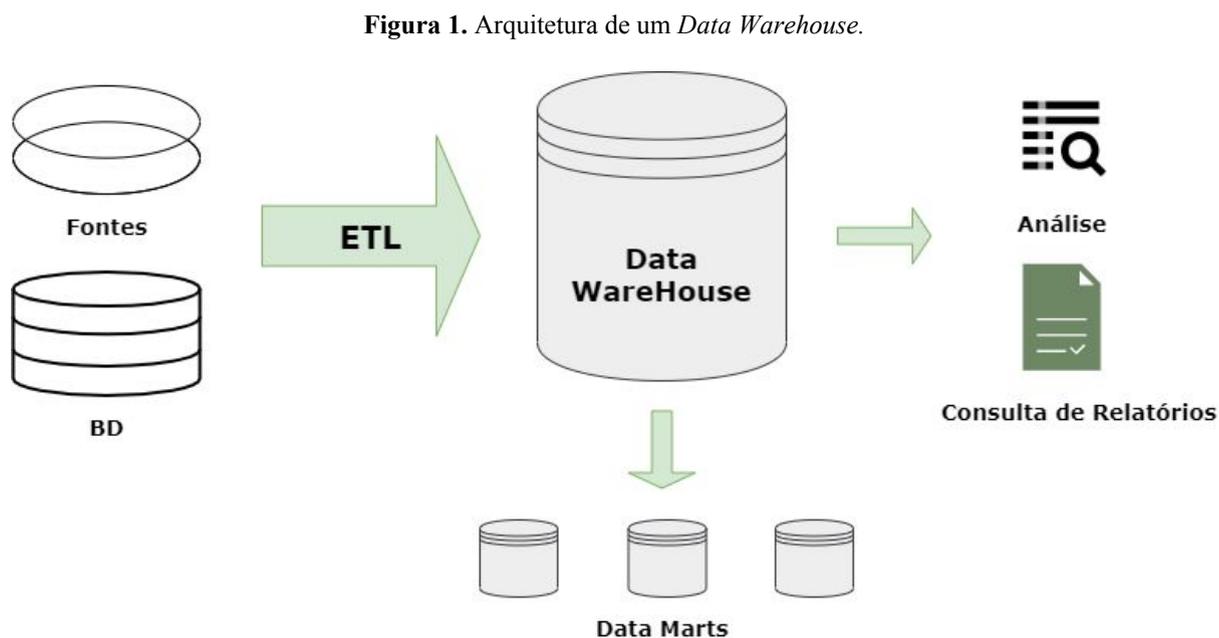
O objetivo deste capítulo é apresentar os principais conceitos e definições que estão intimamente relacionados com o tema proposto. Nas seções a seguir serão abordados os seguintes conceitos: *Data Warehouse*, Processo *ETL*.

2.1 DATA WAREHOUSE

O *Data Warehouse* é um grande banco de dados que armazena dados de múltiplas fontes para o desenvolvimento de informações integradas, com base nos dados de uma organização inteira. Conforme Shams (2001), o DW é “uma plataforma que contém todos os dados da organização, centralizados e organizados, de forma que os usuários possam extrair relatórios analíticos complexos, contendo informações gerenciais para apoio à decisão”.

Segundo Date (2004), os *data warehouses* surgiram pela necessidade de existir uma origem única de dados, limpa e consistente para apoio à decisão e para que isso seja feito sem causar impacto sobre os sistemas operacionais.

O *Data Warehouse* deve corresponder a consultas avançadas de maneira rápida, mostrando detalhes relevantes à resposta. Diante disso, sua arquitetura deve permitir coletar, manipular e apresentar os dados de forma eficiente e rápida. A figura 1 mostra os principais componentes da arquitetura de um *Data Warehouse*.



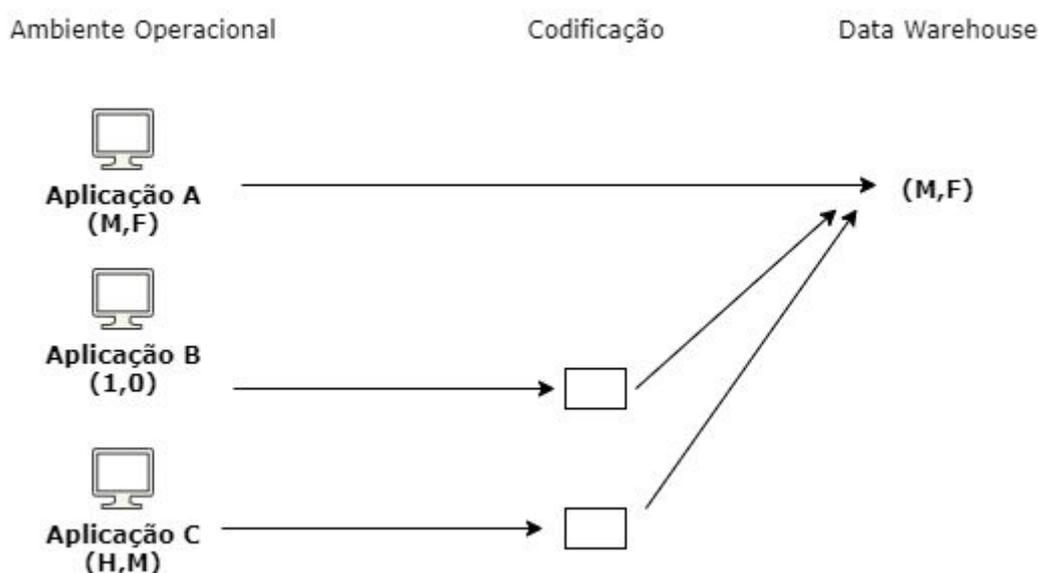
Fonte: Adaptado de Machado (2004)

A arquitetura de um DW contém ferramentas para extrair dados de múltiplas bases de dados operacionais e fontes externas; limpar, transformar e integrar estes dados, carregá-los até o *Data Warehouse* e frequentemente fazer o *refresh*, isto é, promover as atualizações ocorridas nas múltiplas bases de dados operacionais. Em junção ao *Data Warehouse*, pode haver vários *Data Warehouses* departamentais, que são nomeados como *Data Marts*.

Segundo Inmon (1997), um DW deve ser orientado por assuntos, integrado, variável no tempo e não volátil. A seguir estas características serão detalhadas e ilustradas.

- **Orientado por assuntos:** de acordo com Assunção (2001), refere-se ao fato de o *Data Warehouse* armazenar informações agregadas sobre temas específicos, em torno das aplicações da empresa como, clientes, vendas, produtos, etc., enquanto os sistemas tradicionais são orientados a processos como estoques, compras e vendas, etc.
- **Integrado:** segundo Colaço Júnior (2004, p. 16), o *Data Warehouse* deve consolidar dados de diversas origens, o que geralmente envolve diferentes codificações. Os dados devem ser perfeitamente integrados para que ao serem armazenados assumam uma única convenção. A Figura 2, apresenta a idealização do conceito.

Figura 2. Idealização do conceito de Integração.



Fonte: Adaptado de INMON (1997)

A figura 2 demonstra um exemplo de vários sistemas onde os desenvolvedores definiram variadas formas para atribuir o campo SEXO. A “aplicação A” usar o sexo como

“M” e “F”, a “aplicação B”, como 1 e 0 e a “aplicação C”, como “H” e “M”. Entretanto quando SEXO é carregado para o *Data Warehouse*, o mesmo deve ser convertido para um único formato.

- **Variável no tempo:** segundo Inmon (1996), os *Data Warehouses* são variáveis em relação ao tempo, isso nada mais é do que manter o histórico dos dados durante um período muito superior ao dos sistemas transacionais.
- **Não volátil:** no ambiente de um *data warehouse* não é possível fazer modificações nos dados tais como inclusão, alteração ou exclusão, pois após os dados serem carregados, na carga inicial, estes ficam disponíveis apenas para consulta, havendo permissão apenas para a inclusão de mais dados, sem alterar os já existentes.

2.1.1 MODELO DOS DADOS

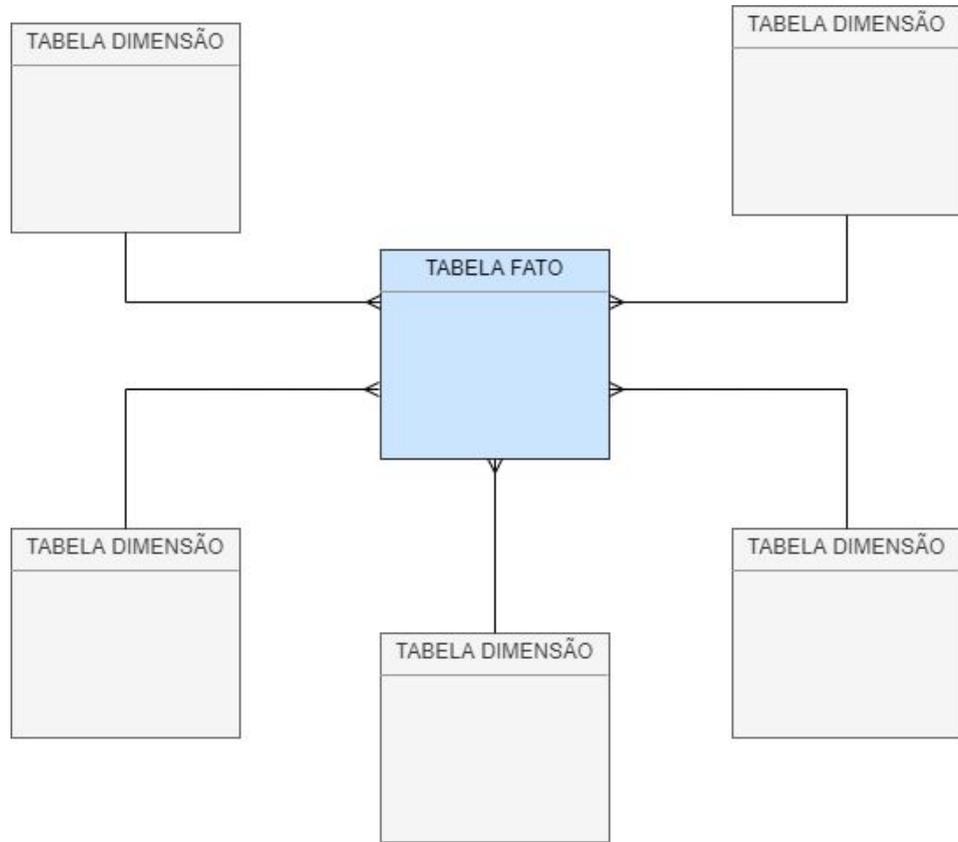
O modelo de dados representa um fator fundamental para o desenvolvimento interativo do *Data Warehouse*. Segundo Musicante (2001, p. 57) “quando os esforços de desenvolvimento são baseados em um único modelo de dados, sempre que for necessário unir estes esforços, os níveis de sobreposição e desenvolvimento desconexo serão muito baixos, pois todos os componentes do sistema estarão utilizando a mesma estrutura de dados”.

De acordo com Assunção (2001) a modelagem de dados consiste na elaboração de um banco de dados usando modelos já consagrados. Nesse sentido, destacam-se dois modelos: o Modelo Entidade-Relacionamento (MER) e o Modelo Dimensional (MD).

No modelo MER os dados estão divididos em tabelas que se relacionam entre si, fazendo com que o diagrama se torne mais complexo. Por essa razão, a utilização do modelo dimensional é apropriada para a análise dos dados.

Segundo Kimball (2008) a modelagem dimensional é uma técnica de projeto lógico que busca apresentar os dados em uma estrutura intuitiva e de alto desempenho através da diminuição do grau de normalização. Os esquemas de dados criados com a Modelagem Dimensional recebem o nome de Esquema Estrela ou Esquema em Estrela (KIMBALL, 2008). Este esquema pode ser visualizado na figura 3.

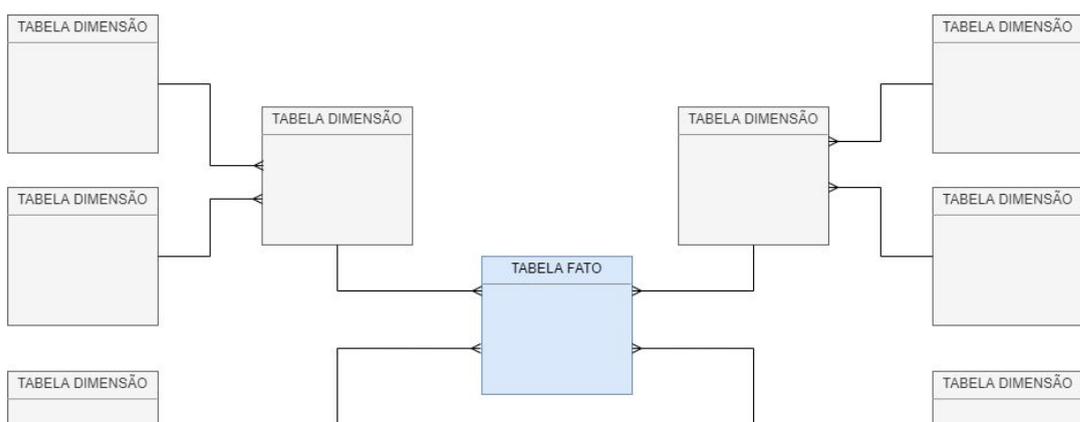
Figura 3. Exemplo de um Esquema Estrela



A figura 3 apresenta um exemplo de um esquema estrela. Este esquema é composto por uma tabela principal localizada ao centro do diagrama, denominada de Tabela de Fatos. Nessa tabela são armazenadas todas as transações e fatos decorrentes dos processos de negócio. A tabela de fatos conecta-se através de junções às demais tabelas auxiliares conhecidas por Dimensões. As dimensões armazenam as características que contextualizam os fatos.

O esquema floco de neve ou *snowflake* é uma extensão do esquema estrela, em que cada uma das “pontas da estrela” passa a ser o centro de outras estrelas, pois cada tabela de dimensão é normalizada, quebrando-se a tabela original ao longo das hierarquias existentes em seus atributos (ANDREATTO, 1999, p. 17). Um exemplo de dimensão de produto pode ser visualizado na figura 4.

Figura 4. Exemplo de um Esquema Floco de Neve



Como pode ser observado na figura 4, a dimensão produto possui sua hierarquia definida e cada um desses relacionamentos gerou uma nova tabela no esquema floco de neve. Segundo Machado (2004), o modelo floco de neve é o resultado da decomposição de uma ou mais dimensões com hierarquias entre seus membros.

2.1.2 DATA MART

“Os *data marts* representam um subconjunto dos DW que permitem o acesso descentralizado à informação. Os DM podem ser direcionados a um departamento ou área específica do negócio” (MACHADO, 2006). Entende-se com isso que *data mart* é um DW de dimensões reduzidas, com custo menor, e tem como objetivo atender apenas uma área do negócio, pois mantém informações sobre um assunto específico que interessa ao departamento envolvido, como vendas.

Conforme Assunção (2001), um *data mart* é um pequeno *Data Warehouse*, que oferece suporte à decisão a um departamento ou um grupo de pessoas, isto é, como um depósito de dados voltado a uma determinada área do negócio. Ele diferencia-se do *Data Warehouse* apenas quanto ao tamanho e ao escopo do problema a ser resolvido. Existem diferentes alternativas de se implementar os *Data Marts*, porém a proposta original é aquela onde os *Data Marts* são desenvolvidos a partir de um DW central.

A Tabela 1 demonstra um breve comparativo entre *Data Mart* e *Data Warehouse*.

Tabela 1. Data Mart x Data Warehouse

<i>Data Mart</i>	<i>Data Warehouse</i>
Departamental	Corporativo
Alto nível de granularidade	Baixo nível de granularidade
Pequena quantidade de dados históricos	Grande quantidade de dados históricos
Tecnologia otimizada para acesso e análise	Tecnologia otimizada para armazenamento e gerenciamento de grandes volumes de dados
Cada departamento tem uma estrutura diferente	Estrutura se adapta ao entendimento corporativo dos dados

Fonte: Inmon (1999)

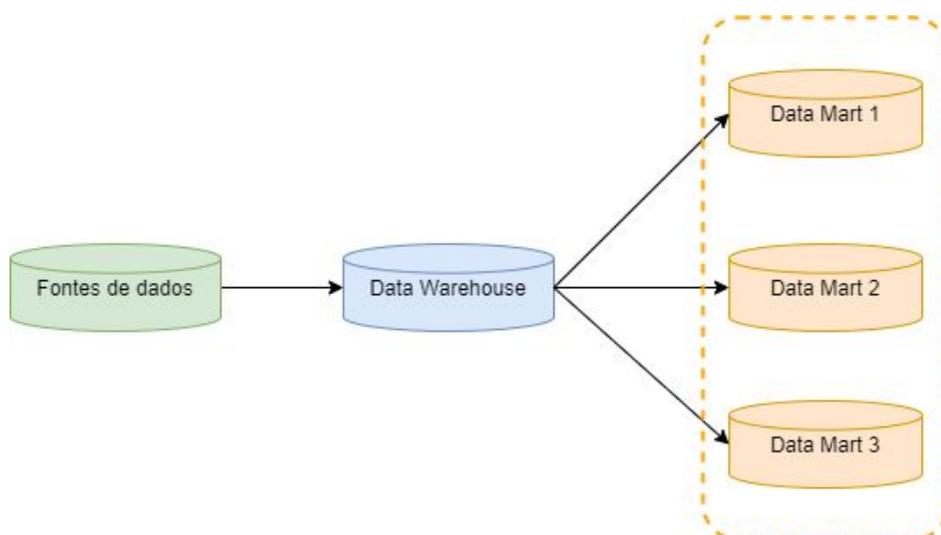
A comparação apresentada pela Tabela 1 tem relação ao tamanho e ao escopo do problema a ser resolvido. Enquanto um *Data Mart* trata de problema departamental ou local,

um *Data Warehouse* envolve o esforço de toda a companhia para que o suporte às decisões atuem em todos os níveis da organização.

2.1.3 IMPLEMENTAÇÃO

Existem três formas que podem ser utilizadas para implementar sistemas de *data warehouse*: *top down*, *bottom up* e a combinação das duas. Ao escolher qual forma de implementação adotar, deve-se levar em consideração os recursos disponíveis, a escolha da arquitetura, as necessidades de acesso a dados, entre outras variáveis. A figura 5 apresenta o desenvolvimento *top down*.

Figura 5. Implementação top down

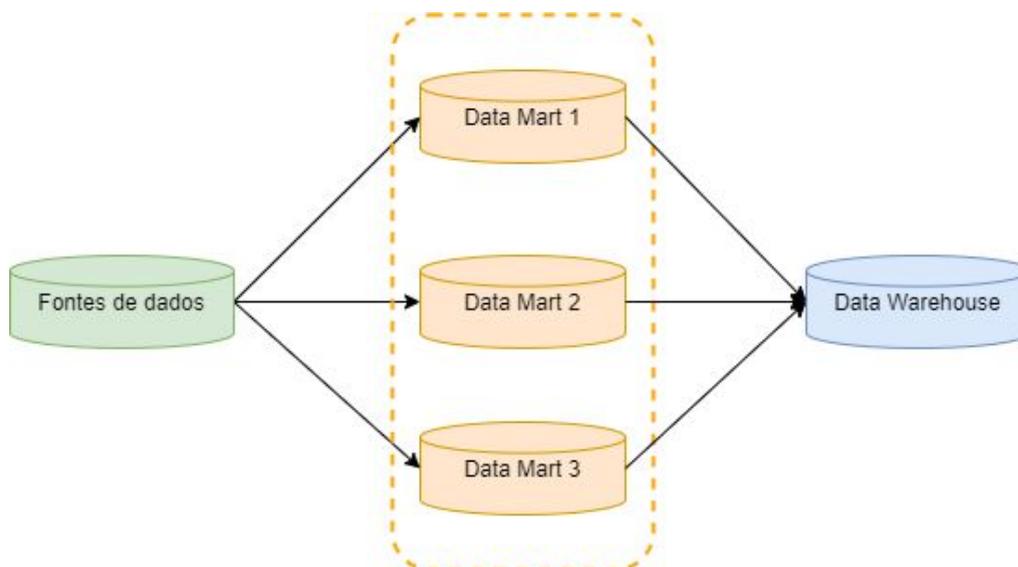


Fonte: Adaptado de Vilarino (2017)

Como pode ser observado na Figura 5, na implementação *top down*, os Data Marts são derivados do *Data Warehouse*. Esta forma de implementação requer a definição do modelo de dados, uma vez que antecede a implementação, bem como necessita de uma visão geral de toda a organização (MAZZOLA, 2002).

A abordagem *top down* tem sido pouco utilizada por necessitar de um Sistema de Informação centralizador que gerencie todos os recursos de hardware, aumentando seu tempo de implementação, além de dificultar e tornar maiores seus custos (MAZZOLA, 2002). A figura 6 apresenta o desenvolvimento *bottom up*.

Figura 6. Implementação bottom up



Fonte: Adaptado de Vilarino (2017)

Na Figura 6, o desenvolvimento *bottom up* do *Data Warehouse* é derivado dos *Data Marts*. Esta forma de implementação apresenta resultados imediatos, tendo desta forma um retorno do investimento mais rápido, se comparado ao *top down* (MAZZOLA, 2002).

Enfim, a última forma de implementação consiste na combinação das técnicas *top down* e *bottom up*, mesclando a maior quantidade de características positivas de cada uma das técnicas. Sendo assim, requer um cuidadoso monitoramento do processo, bem como acompanhamento e gerenciamento de novos possíveis requerimentos (MAZZOLA, 2002).

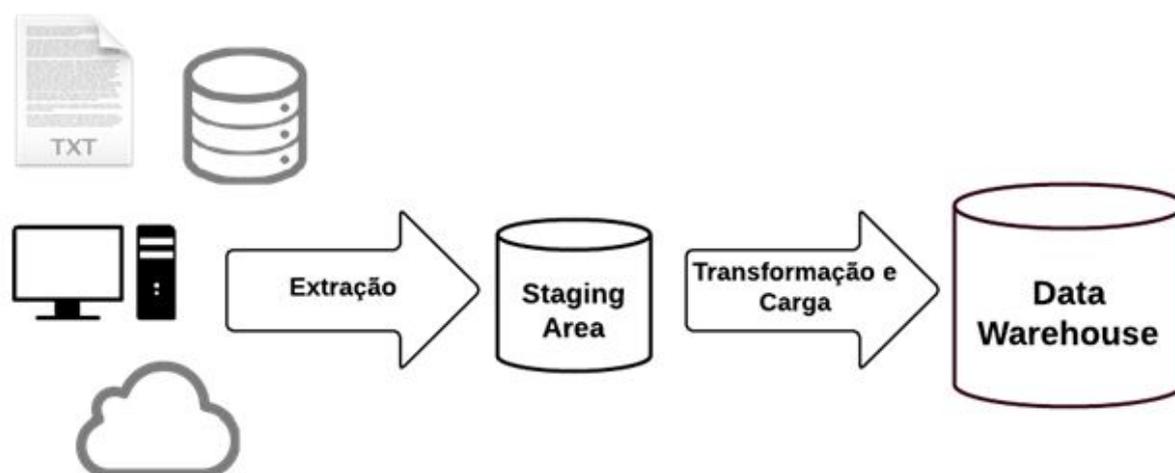
2.2 PROCESSO ETL

Essa fase é responsável pela preparação dos dados que farão parte do DW ou DM. Ela se dá basicamente em três passos: extração, transformação e carga de dados.

Segundo Guerra (2005), a etapa de *ETL* é uma das mais críticas de um projeto de DW, pois uma informação carregada erroneamente trará consequências imprevisíveis nas fases posteriores. O objetivo desta fase é fazer a integração de informações de fontes múltiplas e complexas.

Basicamente, divide-se esta etapa em três passos: extração, transformação e carga dos dados, bem como demonstrado na figura 7.

Figura 7. ETL



FONTE: CANALTECH, 2014.

As três etapas (Extração, Transformação e Carga) serão mais detalhadas a seguir.

- Extração

Para se obter dados para o ambiente de *Data Warehouse*, deve-se iniciar o processo pela extração de dados. Na etapa de Extração do processo *ETL*, ocorre a extração dos dados de origem. Esses dados podem ser provenientes de diversas fontes como os SGBDs (Sistemas de Gerenciamento de Banco de Dados), planilhas eletrônicas, arquivos textos, entre outros.

- Transformação

A transformação dos dados é a fase subsequente à sua extração. Esta fase não só transforma os dados, mas também realiza a limpeza deles. A correção de erros de digitação, a descoberta de violações de integridade, a substituição de caracteres desconhecidos, as padronizações de abreviações podem ser exemplos dessa limpeza (GONÇALVES, 2003). O objetivo principal da limpeza dos dados é a redução dos erros, melhorando suas qualidades e fornecendo dados úteis ao usuário.

- Carga

A última etapa do processo de *ETL* é a carga. Os dados já extraídos e transformados são carregados para dentro do DW. Os dados a serem entregues ao *data warehouse* são as tabelas de dimensões e tabelas de fatos.

Apesar de existirem ferramentas de *ETL* ainda se tem a necessidade de criar rotinas de cargas para atender determinadas situações que poderão ocorrer. O que se pode afirmar sobre

os benefícios do uso das ferramentas de *ETL* é que os benefícios serão bastante vistosos e a produtividade aumentará consideravelmente. (PRIMAK, 2008).

3 METODOLOGIA

Neste capítulo são apresentadas as ferramentas utilizadas e a metodologia adotada para elaboração do presente trabalho.

3.1 FERRAMENTAS

As ferramentas utilizadas para o desenvolvimento do *Data Mart* foram o SGBD (Sistema de Gerenciamento de Banco de Dados) *PostgreSQL*, *Pentaho Data Integration – PDI* e o sistema *E-ext* do Ministério Público do Estado do Tocantins, os quais serão ilustrados na figura 8 em que demonstra como os softwares estão relacionados.

Figura 8. Relacionamento dos Softwares



A relação entre as ferramentas, mostradas na Figura 8, corresponde a base de dados *E-ext*, que foram os dados processados no projeto, e utilizando a ferramenta *Pentaho Data Integration* esses dados foram extraídos das suas origens. O próprio PDI realizou o processo de extração, transformação e carga dos dados no *data mart* construído através da ferramenta *PostgreSQL*. Todos estes softwares utilizados serão detalhados nas próximas seções.

3.1.1 E-ext

O Sistema de Procedimentos Extrajudiciais (*E-ext*) foi desenvolvido pelos analistas do Departamento de Tecnologia da Informação do MPE-TO que visa gerenciar a tramitação de procedimentos, padronizar o tratamento de dados, além de fornecer informações necessárias aos órgãos de supervisão do Ministério Público. O sistema considera procedimentos como o conjunto de documentos ou eventos produzidos e registrados na plataforma.

A ferramenta permite a comunicação a distância por meio de documentos ou arquivos digitais, na qual o usuário digitaliza documentos originalmente físicos e produz para o meio digital. A disponibilização é feita de forma que a identificação inequívoca do signatário é realizada através de assinatura digital e assinatura eletrônica.

O sistema permite aos usuários produzir documentos, anexar arquivos no formato PDF, além de outras mídias de formatos abertos para áudio, vídeos e imagens. No momento atual, o sistema *E-ext* soma 25.236 procedimentos, contendo 263.607 documentos de procedimentos.

Os usuários do *E-ext* são internos e externos, em que os internos são servidores, voluntários e estagiários do Ministério Público e os externos são partes, procuradores, advogados, defensores e interessados na relação procedimental. As partes são as pessoas envolvidas no processo.

3.1.2 Pentaho Data Integration

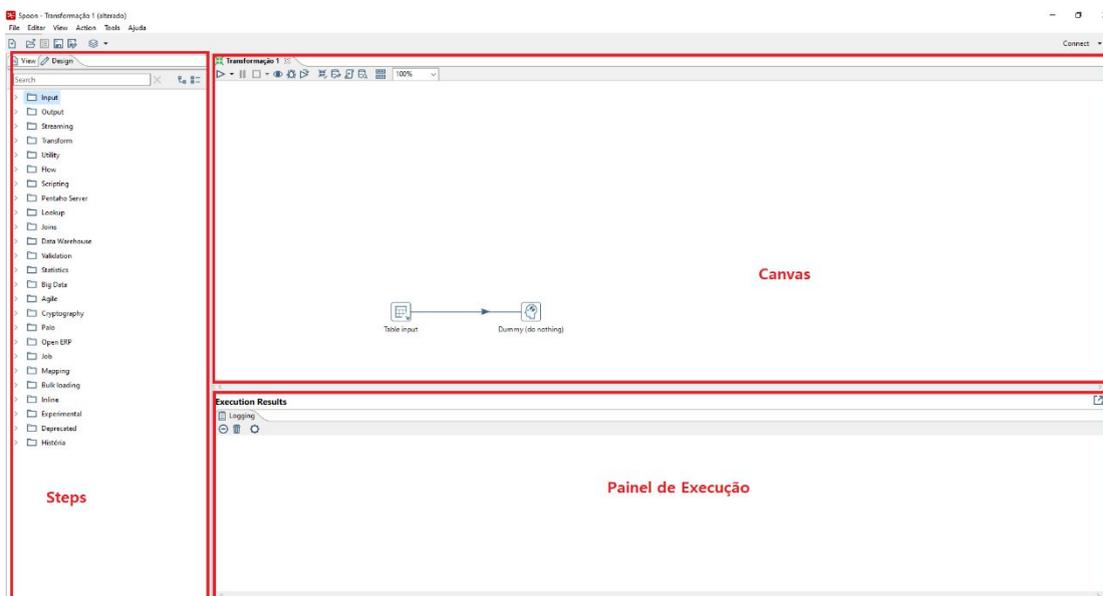
O *Pentaho Data Integration* (PDI), também conhecido como *Kettle*, é um poderoso mecanismo usado para criar processos de extração, transformação e carga (*ETL*) que alimentam o banco de dados (PENTAHO, 2020). O mecanismo fornece a migração de dados entre aplicações ou base de dados, exportação de dados e integração de aplicações.

Segundo Bouman (2009), a arquitetura do *Pentaho Data Integration* é baseada na linguagem Java, e consiste em quatro componentes básicos:

- **Spoon:** ferramenta de modelagem gráfica direcionada ao usuário, onde se define a entrada, transformações e saída de dados;
- **Pan:** aplicativo de linha de comando para executar as transformações feitas no Spoon;
- **Chef:** ferramenta de modelagem gráfica direcionada para criação de *Jobs*, que consiste em tarefas como transformação, downloads, etc., no qual são colocados em um fluxo de controle;
- **Kitchen:** aplicativo de linha de comando para executar os *Jobs* criados no Chef.

O *Spoon* é a interface gráfica do PDI que facilita na concepção de rotinas e lógica *ETL*. A seguir, apresentamos a interface do *Spoon*.

Figura 9. Interface gráfica do Spoon



A interface do *Spoon*, apresentada na figura 9, permite criação e projeção de transformações e tarefas *ETL*, além de agendar quando as tarefas devem ser executadas. A marcação *Step* indica os blocos de construção de uma transformação, por exemplo, uma entrada de arquivo de texto ou uma saída de tabela. A indicação *Canvas* corresponde a área utilizada para projetar e construir transformações e tarefas para as atividades de *ETL*. Por último, a marcação *Painel de Execução* contém várias guias diferentes que ajudam a ver como a transformação foi executada, erros de identificação e desempenho do monitor.

A ferramenta apresentada foi utilizada para automatizar o processo *ETL* por meio dos *Jobs* criados na ferramenta. Os *Jobs* são usados para coordenar atividades *ETL*, como por exemplos; definindo o fluxo e as dependências para quais transformações de ordem devem ser executadas; executar operações de banco de dados de carregamento em massa. Além disso, é possível agendar o trabalho para ser executado.

3.1.3 PostgreSQL

É um sistema de banco de dados relacional e de código aberto que usa e estende a linguagem SQL combinada com muitos recursos que armazenam e escalam com segurança as cargas de trabalho de dados mais complicadas (POSTGRESQL,2020). Neste trabalho, o *PostgreSQL* foi utilizado para construir o *data mart* e armazenar a carga de dados executada no processo ETL.

3.2 MÉTODOS

Este projeto de pesquisa foi desenvolvido em etapas, conforme ilustrado na Figura 9.

Figura 10. Metodologia



As cinco etapas ilustradas na Figura 9 demonstram o processo aplicado para o desenvolvimento do *data mart* e automatização do processo *ETL* que este trabalho propõe.

Entendimento do modelo ER do E-ext foi a primeira etapa realizada no desenvolvimento deste projeto de pesquisa. Esta etapa consistiu em um estudo dos dados armazenados no MPE, a fim de extrair o entendimento necessário para a criação de um modelo abstrato do *data mart*.

A próxima etapa realizou-se a **Definição do modelo abstrato do data mart**, onde foram definidas as dimensões, os fatos, as métricas e modelo estrela para a constituição do *data mart*.

Finalizada a etapa de **Definição do modelo abstrato do data mart**, foi iniciada a **Construção do modelo físico do data mart**, utilizando o *PostgreSQL* para implementar uma estrutura do modelo físico, para armazenamento dos dados posteriores a esta etapa.

Com o modelo físico pronto, foi iniciada a etapa de **Integração dos dados**. Esta etapa consistiu em utilizar a ferramenta *Pentaho Data Integration* para integração de dados. O processo foi dividido em duas etapas: integração dos dados para criação das dimensões e integração dos dados para criação da tabela fato.

Por fim, na etapa de **Criação de jobs** realizou-se a execução da automatização do processo *ETL*, que foi executado após a criação do fato e das dimensões, utilizando a ferramenta *Pentaho Data Integration*, com a criação de *jobs*.

4 RESULTADOS E DISCUSSÃO

O intuito deste trabalho foi construir um *data mart* e automatizar o processo *ETL* para fornecer uma estrutura centralizada para acesso aos dados do sistema de procedimento extrajudiciais do MPE/TO (e-ext). Diante disso, esta seção apresenta os resultados do desenvolvimento do *data mart* utilizando a ferramenta PostgreSQL, bem como a automatização do processo *ETL* utilizando a ferramenta *Pentaho Data Integration – PDI* no *Data Mart* construído.

4.1 PROJETO DO MODELO LÓGICO

Através das análises realizadas e da obtenção de um conhecimento mais aprofundado sobre os dados, foram levantadas necessidades que direcionaram a criação do modelo lógico do *Data Mart*, mediante os seguintes questionamentos:

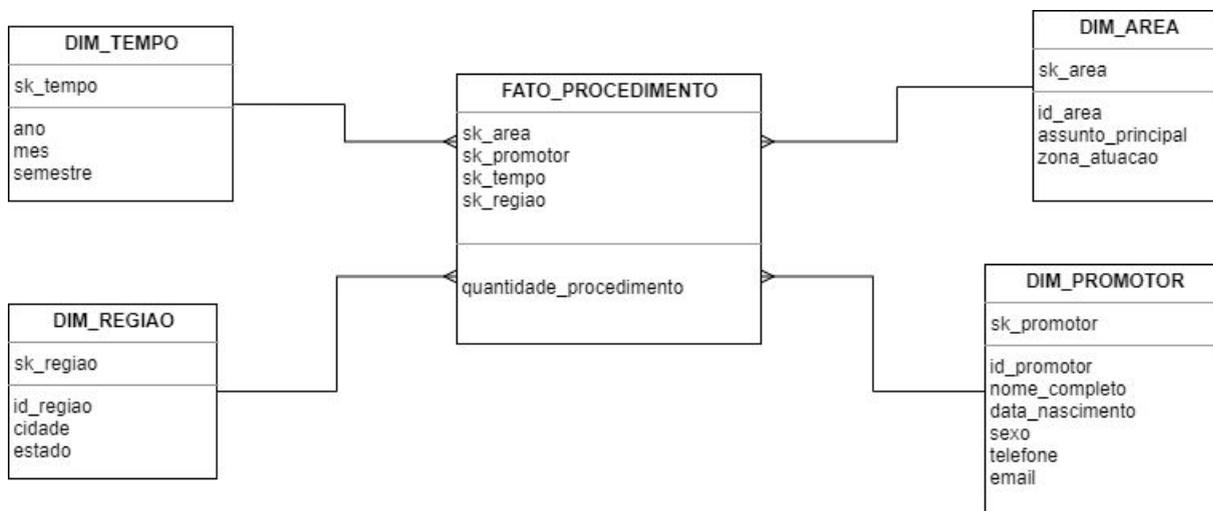
1. Em quantos procedimentos extrajudiciais da área X o promotor Y participou?
2. No ano de 2019 a área criminal teve a instauração de quantos procedimentos em que o promotor X atuou?
3. Qual o percentual de procedimentos extrajudiciais realizados na área eleitoral no primeiro semestre do ano de 2020 na cidade de Araguaína?
4. Qual o índice de crescimento de procedimentos extrajudiciais de determinada área X, nas cidades do estado Tocantins desde 2017?

Com base nessas necessidades, foi possível elaborar o modelo multidimensional do *data mart*. A abordagem utilizada para implementação do projeto foi a modelo estrela, de forma que a escolha desse modelo multidimensional se deu em virtude do desempenho do modelo estrela ser superior em 30% ao modelo *SnowFlake* (MACHADO,2000).

A partir da escolha do modelo, foi identificada a tabela fato e as demais tabelas dimensões do modelo estrela. A tabela fato refere-se a tudo aquilo que pode ser mensurado no questionamento, sendo o fato **procedimento** apresentando-se em destaque. Já as tabelas dimensões, que complementam o fato, foram identificadas como **área**, **região**, **tempo** e **promotor**.

O modelo criado para o *Data Mart* (Figura 11), consiste na tabela “FATO PROCEDIMENTO” ao centro e ao seu redor estão dispostas as dimensões “DIM_AREA”, “DIM_REGIAO”, “DIM_TEMPO” e “DIM_PROMOTOR”.

Figura 11. Modelo logico do data mart.

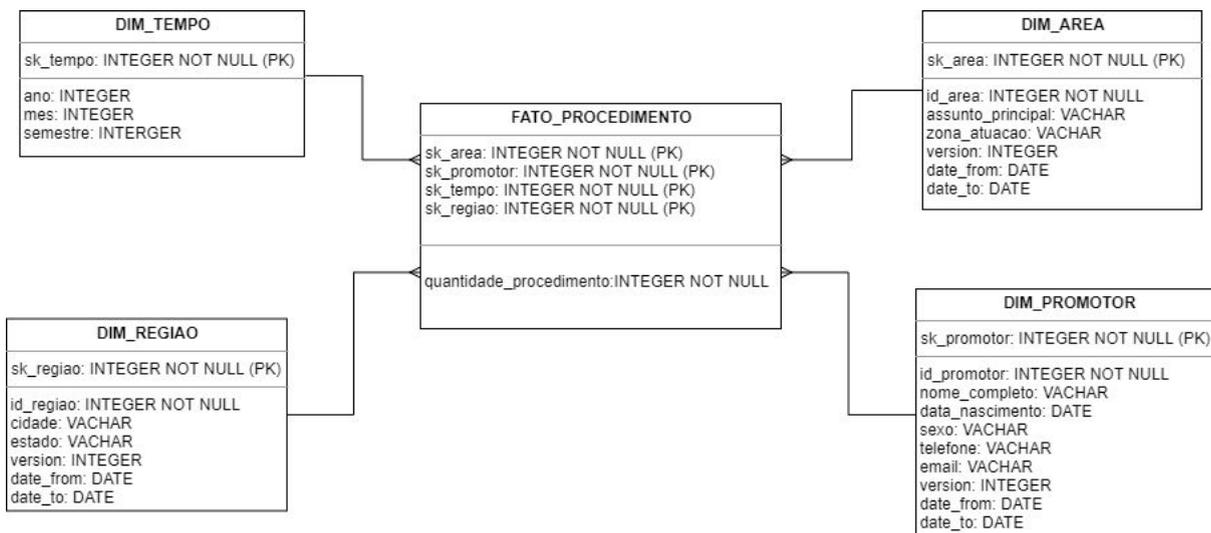


Para cada uma das tabelas de dimensões foi criada uma chave substituta, essa chave representa um campo com as características da chave primária e é obtida através de um número sequencial gerado. O uso deste tipo de chave permite tratar os dados históricos na tabela de fato e as modificações dos dados nas dimensões. No caso de alteração em algum dos dados da dimensão é inserido um novo registro, com uma nova chave primária e com os dados alterados.

4.2 PROJETO DO MODELO FÍSICO

O modelo lógico elaborado forneceu uma base para construção do modelo físico. O modelo físico apresentado na figura 12, ajuda a visualizar como a estrutura do banco de dados foi implementada, evidenciando a relação entre tabela fato e as tabelas dimensões do *data mart*.

Figura 12. Modelo físico do data mart.



O modelo implementado, Figura 12, indica o fato, suas dimensões e seus campos, apresentados em detalhes a seguir nas tabelas 2 a 6.

Tabela 2 – Fato Procedimento

FATO_PROCEDIMENTO		
Campo	Tipo de Dados	Descrição
sk_promotor	integer	Chave substituta; representa o identificador da dimensão dim_promotor
sk_area	integer	Chave substituta; representa o identificador da dimensão dim_area
sk_regiao	integer	Chave substituta; representa o identificador da dimensão dim_regiao
sk_tempo	integer	Chave substituta; representa o identificador da dimensão dim_tempo
quantidade_procedimento	integer	Campo; contém informação modificada a partir das dimensões

A tabela 2 é responsável por armazenar todos os atributos identificadores das tabelas de dimensões, como código identificador da área, região, tempo e promotor, além da mesma possuir o campo que vai permitir mensurar a quantidade de procedimentos.

Tabela 3 – Dimensão Promotor

DIM_PROMOTOR

Campo	Tipo de Dados	Descrição
sk_pessoa	integer	Chave substituta; representa o identificador da dimensão
id_promotor	integer	Chave primária; representa o identificador do sistema de origem (<i>E-ext</i>)
nome_completo	varchar	Campo; contém informações baseadas nos dados do sistema de origem(<i>E-ext</i>).
data_nascimento	date	Campo; contém informações baseadas nos dados do sistema de origem(<i>E-ext</i>).
sexo	varchar	Campo; contém informações baseadas nos dados do sistema de origem(<i>E-ext</i>).
telefone	varchar	Campo; contém informações baseadas nos dados do sistema de origem(<i>E-ext</i>).
email	varchar	Campo; contém informações baseadas nos dados do sistema de origem(<i>E-ext</i>).
version	integer	Campo; representa a versão da entrada da dimensão promotor
date_from	date	Campo; representa a data de início da validade da carga dimensão
date_to	date	Campo; representa a data final da validade da carga dimensão

A tabela 3 contém as informações de usuários promotor conforme o sistema *E-ext*, sendo composta por uma chave de relacionamento e um código identificador da tabela, além disso tem os campos nome completo, data nascimento, sexo, estado civil, telefone e e-mail. Vale ressaltar que foram adicionados os campos version, date_to e date_from para controle do rastreamento e armazenamento de modificações no registro da tabela.

Tabela 4 – Dimensão Regiao

DIM_REGIAO		
Campo	Tipo de Dados	Descrição
sk_regiao	integer	Chave substituta; representa o identificador da dimensão
id_regiao	integer	Chave primária; representa o identificador do sistema de origem (<i>E-ext</i>)
cidade	varchar	Campo; contém informações baseadas nos dados do sistema de origem(<i>E-ext</i>).

estado	varchar	Campo; contém informações baseadas nos dados do sistema de origem(<i>E-ext</i>).
version	integer	Campo; representa a versão da entrada da dimensão região
date_from	date	Campo; representa a data de início da validade da carga dimensão
date_to	date	Campo; representa a data final da validade da carga dimensão

A tabela 4 contém as informações da cidade e estado de tal procedimento conforme o sistema *E-ext*, sendo composta uma chave de relacionamento e um código identificador da tabela e campos ano, mês e semestre. Vale ressaltar que foram adicionados os campos version, date_to e date_from para controle do rastreamento e armazenamento de modificações no registro da tabela.

Tabela 5 – Dimensão Area

DIM_AREA		
Campo	Tipo de Dados	Descrição
sk_area	integer	Chave substituta; representa o identificador da dimensão
id_area	integer	Chave primária; representa o identificador do sistema de origem (<i>E-ext</i>)
assunto_principal	varchar	Campo; contém informações baseadas nos dados do sistema de origem(<i>E-ext</i>).
zona_atuacao	varchar	Campo; contém informações baseadas nos dados do sistema de origem(<i>E-ext</i>).
version	integer	Campo; representa a versão da entrada da dimensão área
date_from	date	Campo; representa a data de início da validade da carga dimensão
date_to	date	Campo; representa a data final da validade da carga dimensão

A tabela 5 contém informações de uma área específica de tal procedimento conforme o sistema *E-ext*, sendo composta por uma chave de relacionamento e um código identificador

da tabela, além dos campos assunto principal e zona de atuação. Vale ressaltar que foram adicionados os campos *version*, *date_to* e *date_from* para controle do rastreamento e armazenamento de modificações no registro da tabela.

Tabela 6 – Dimensão Tempo

DIM_TEMPO		
Campo	Tipo de Dados	Descrição
sk_tempo	integer	Chave substituta; representa o identificador da dimensão
ano	integer	Campo; contém informações baseadas nos dados do sistema de origem(<i>E-ext</i>).
mes	integer	Campo; contém informações baseadas nos dados do sistema de origem(<i>E-ext</i>).
semestre	integer	Campo; contém informações baseadas nos dados do sistema de origem(<i>E-ext</i>).

A tabela 6 corresponde aos dados do tempo que foi inserido tal procedimento no sistema *E-ext*, sendo composta por código identificador da tabela e os campos ano, mês e semestre.

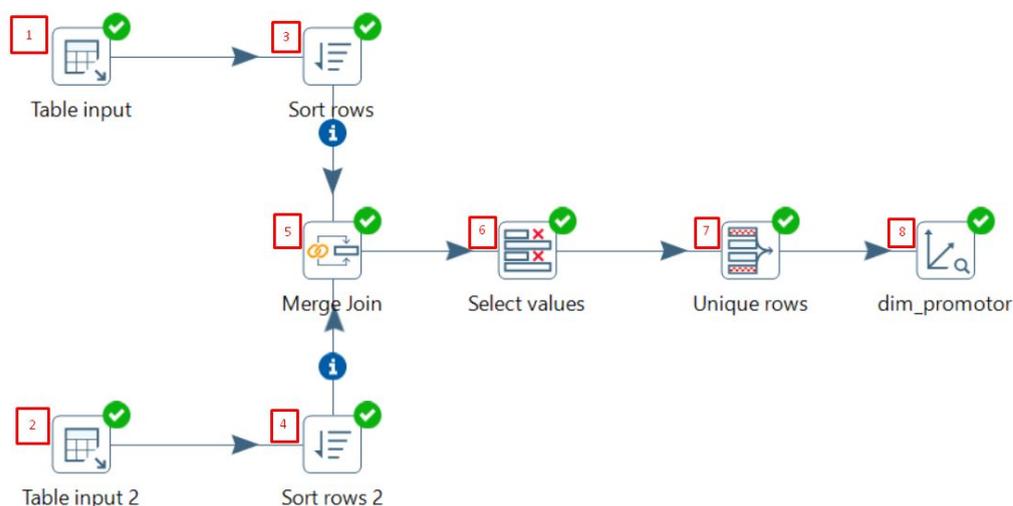
4.3 PROCESSO ETL

Para desenvolver os processos de *ETL* desse projeto, foi utilizado o software *Pentaho Data Integration* – PDI. A realização de um processo de *ETL* na ferramenta consiste na criação de transformações. As transformações são compostas de *steps* responsáveis por realizar a extração, transformação e a carga dos dados no *data mart* construído.

4.3.1 Processo ETL da tabela dimensão (DIM_PROMOTOR)

O processo *ETL* da dimensão Promotor consistiu na criação de transformações no PDI, em que resultou a integração dos dados na dimensão promotor no *data mart*. A figura 13 apresenta o fluxo de transformações da ferramenta PDI para a carga da tabela dimensão promotor do *data mart*.

Figura 13. Processo *ETL* da tabela dimensão promotor.



No fluxo apresentado na figura 13, pode-se destacar 8 marcações que representam as etapas do fluxo de trabalho para a integração de dados da dimensão Promotor.

O *step Table input*, indicação 1, realiza a leitura das informações da tabela Pessoa disponível no banco de dados do sistema *E-ext*, utilizando instruções *SQL* predeterminada, e através de um *hop*, objeto ao qual permite que metadados do esquema passem de um passo para outro, conecta-se com o *step Sort rows*, marcação 3, que classifica as linhas em ordem crescente com base no campo *idPessoa*.

Por outro lado, o *step Table input 2*, marcação 2, efetua a leitura das informações da tabela Promotor disponível no banco de dados secundário do sistema *E-ext*, utilizando instruções *SQL* predeterminada, e através de um *hop* conecta-se com o *step Sort rows 2*, marcação 4, que classifica as linhas em ordem crescente com base no campo *Nome*.

O *step Merge join*, indicação 5, mescla os conjuntos de dados provenientes das etapas 3 e 4. Em seguida, o *step Select values*, a marcação 6 executa ações de selecionar, alterar e remover os dados resultantes da mesclagem da etapa 5. O *step Unique rows*, indicação 7, remove as linhas duplicadas do fluxo de entrada. Por último, *step dim_promotor*, marcação 8, realiza a inserção dos registros na tabela “DIM_PROMOTOR” do *data mart*.

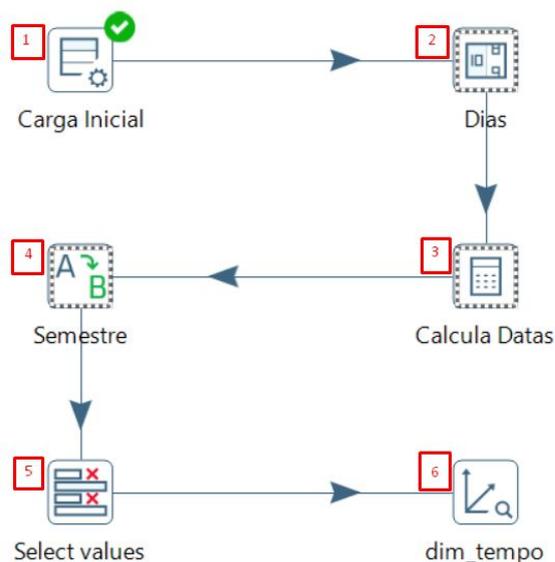
O fluxo de integração da dimensão promotor, apresentado na figura 13, se repetiu para as tabelas dimensões “DIM_AREA” e “DIM_REGIAO”.

4.3.2 Processo ETL da tabela dimensão (DIM_TEMPO)

O processo *ETL* da dimensão Tempo consistiu na criação de transformações no PDI, em que resultou a integração dos dados na dimensão tempo no *data mart*. A figura 14

apresenta o fluxo de transformações da ferramenta PDI para a carga da tabela dimensão tempo do *data mart*.

Figura 14. Processo *ETL* da tabela dimensão tempo.



No fluxo apresentado na figura 14, pode-se destacar 6 marcações que representam as etapas do fluxo de trabalho para a integração de dados da dimensão tempo.

O *step Carga inicial*, indicação 1, é gerado uma saída de 10000 linhas, e por meio de um *hop* é conectado ao *Step Dias*, marcação 2, que adiciona uma sequência ao fluxo e conecta-se o *step Calcula Datas*, marcação 3, que executa funções predefinidas nos valores dos campos de entrada para calcular uma data, ano, mês, dia da semana, dia do mês e dia do ano.

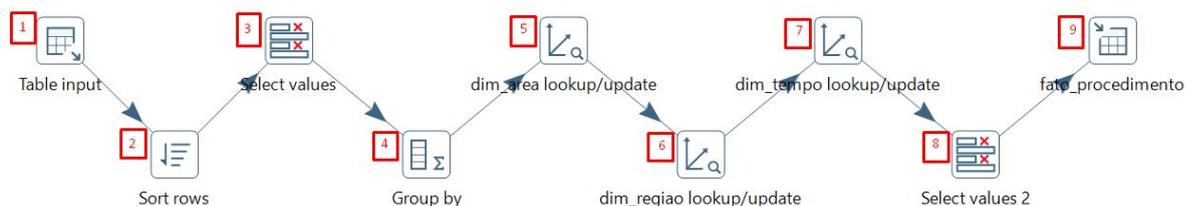
O *step Semestre*, marcação 4, recebe os valores do *step Calcula data* e atribui intervalos com base nos meses calculados para se obter o campo semestre, e conecta-se ao *step Select values*, marcação 5, que seleciona os valores resultantes das etapas de marcação 3 e 4. Por último, o *step dim_tempo*, marcação 6, realiza a inserção dos registros na tabela “DIM_TEMPO” do *data mart*.

4.3.3 Carga da tabela fato (FATO_PROCEDIMENTO)

O processo de integração do fato Procedimento é a criação de um processo *ETL* utilizando o PDI para integrar os dados resultando das dimensões área, região, promotor e

tempo. A figura 15 apresenta o fluxo de transformações da ferramenta PDI para a carga da tabela fato procedimento do *data mart*.

Figura 15. Processo *ETL* da tabela fato procedimento.



No fluxo apresentado na Figura 15, pode-se destacar 9 marcações que representam as etapas do fluxo de trabalho para a integração de dados da dimensão tempo.

O *step Table input*, indicação 1, é realizado a leitura das informações da tabela Promotor do *data mart*, utilizando instruções *SQL* predeterminada, e através de um *hop*, conecta-se com o *step Sort rows*, marcação 2, que classifica as linhas em ordem crescente com base no campo *idPromotor*.

Em seguida, o *step Select values*, marcação 3, executa ações de selecionar, alterar e remover os dados resultantes da etapa 1, e conecta-se ao *step Group by*, marcação 4, que permite agrupar-se linhas com base no campo *IdPromotor* para gerar valores agregados para o grupo, sendo possível contar o número de procedimentos.

Logo após, o *step dim_area lookup/update*, indicação 5, consiste em consultar e atualizar valores na tabela “DIM_AREA” no *data mart*, e os resultados obtidos foram carregados por meio de um *hop* para o *step dim_regiao lookup/update*. Esta etapa, marcação 6, também consiste em consultar e atualizar valores na tabela “DIM_REGIAO” no *data mart*, e os resultados obtidos foram carregados por meio de um *hop* para o *step dim_tempo lookup/update*. Esta etapa, marcação 7, também equivale em consultar e atualizar valores na tabela “DIM_TEMPO” na *data mart*.

Por último, o *step select values 2*, marcação 8, executa ações de selecionar, alterar e remover os dados resultantes das etapas anteriores e conecta-se ao *step fato_procedimento*, marcação 9, no qual, realizar a inserção dos registros na tabela “FATO_PUBLICACAO” no *data mart*.

4.3.2 Execução das transformações do *Data Mart*

Para a execução das transformações foi criado um *job* utilizando o PDI, que executa sequencialmente as transformações podendo ser agendado um horário para execução automaticamente. Normalmente, é agendado no sistema operacional para que seja executado em horários fora do expediente da organização, visando não comprometer o desempenho do sistema transacional. No projeto em questão, foi utilizado o agendador de tarefas do windows, no qual foi criado uma tarefa que executasse o *job* todos os dias às 22:00 horas. O fluxo demonstrado na figura 16 é realizado a automatização do processo de cargas das dimensões e do fato com *job*.

Figura 16. Jobs



No fluxo apresentado na figura 16, pode-se destacar 7 marcações que representam as etapas do fluxo de trabalho para a integração dos processos *ETL* das dimensões de Tempo, Area, Regiao, Promotor e do fato Procedimento.

O *step START*, marcação 1, indicação 1, é utilizado para definir o horário que será executado este *job*, na etapa *step Dim_Promotor*, marcação 2, permite acessar a atividade *ETL* da dimensão promotor, de modo que executar a transformação especificada. Em seguida, o *step Dim_Regiao*, indicação 3, permite acessar a atividade *ETL* da dimensão regiao, de modo que execute a transformação especificada. O *step Dim_Area*, marcação 4, permite acessar a atividade *ETL* da dimensão area, de modo que execute a transformação especificada.

O *step Dim_Tempo*, marcação 5, permite acessar a atividade *ETL* da dimensão tempo, de modo que executar a transformação especificada. Em seguida, o *step Dim_Procedimento*, marcação 6, permite acessar a atividade *ETL* do fato procedimento, de modo que execute a transformação especificada. Por último, o *step Success*, marcação 7, é utilizado para sinalizar o fim da *job*.

5 CONSIDERAÇÕES FINAIS

Neste trabalho realizou-se o desenvolvimento de um *data mart* e a automatização do processo *ETL* dos dados de procedimentos extrajudiciais do MPE/TO, com o objetivo de centralizar os dados referentes ao sistema de procedimentos extrajudiciais (*E-ext*). A ideia inicial desse projeto nasceu pelo motivo dos dados produzidos no sistema *E-ext* serem armazenados em bases de dados distintas, não havendo um padrão na estrutura dos dados. Os conceitos estudados sobre *Data Warehouse* e processo *ETL* promoveram suporte referente a métodos de construir um ambiente de *Data Mart*.

Com a implementação do ambiente de *Data Mart* apresentado, pôde-se constatar uma nova maneira de estruturação dos dados, permitindo o acesso aos dados de forma intuitiva e com altos índices de performance na extração de informações de interesse do órgão.

No desenvolvimento do trabalho se fez o uso da ferramenta *Pentaho Data Integration*, por sua interatividade, rapidez e precisão, e porque possibilita integrar grandes volumes de dados. Foram necessários alguns estudos sobre as ferramentas utilizadas, com o intuito de obter maior conhecimento sobre sua importância. O desenvolvimento iniciou-se com o entendimento dos dados, que serviram como base para projetar a modelagem e construir o *data mart*. Com a estrutura implementada foram iniciados os processos de integração dos dados no *data mart* e por último foram realizadas a automatização desses processos através da criação de *jobs*.

Dentre os desafios para o construir o *data mart*, o principal foi o entendimento da regra de negócio do sistema *e-ext*, pois se trata de dados específicos voltado a processos judiciais.

Como proposta para trabalhos futuros, pretende-se realizar a criação de outros *Data Marts* voltados aos demais departamentos do Ministério Público do Tocantins, podendo assim estabelecer um só *data warehouse*. Além disso, um estudo detalhado da ferramenta *Pentaho Data Integration* seria válido, isso porque a ferramenta apresenta uma abordagem ampla no quesito processo *ETL*. Outro fator relevante é utilização de ferramentas relacionadas às consultas para o usuário final, como elaborações de relatórios, pesquisas informativas e visualização de dados com base nos dados armazenados no *data mart*.

REFERÊNCIAS

ANDREATTO, Ricardo. Construindo um data warehouse e analisando suas informações com data mining e OLAP. Valinhos (SP), 1999. 46f. Monografia (Graduação em Ciência da Computação) – Faculdade de Ciências Administrativas de Valinhos.

ASSUNÇÃO, Adriano Lopes. DATA WAREHOUSE: MERCADO DE TRABALHO PARA O GESTOR DA INFORMAÇÃO? 2001. 57 f. TCC (Graduação) - Curso de Gestão da Informação, Universidade Federal do Paraná, Curitiba, 2001.

BISPO, Carlos Alberto Ferreira. Data warehouse. In: _____. Uma análise da nova geração de sistemas de apoio à decisão. São Carlos (SP), 1998. 23f. Dissertação (Mestrado em Engenharia da Produção) – Escola de Engenharia de São Carlos, Universidade de São Paulo.

BOUMAN, R. Roland. Pentaho Data Integration: Kettle turns data into business. Roland Bouman's blog. 2009. Disponível em: <http://rpbouman.blogspot.com.br/2006/06/pentahodata-integration-kettle-turns.html>. Acesso em: 28 de abr. 2020.

COLAÇO JÚNIOR, M. Projetando Sistemas de Apoio à Decisão Baseados em Data Warehouse. Rio de Janeiro: Axcel Books, 2004.

FERREIRA, Adriano. Tecnologia na contabilidade: análise de Big Data. Disponível em: <<https://www.dominiosistemas.com.br/blog/tecnologia-na-contabilidade-analise-de-big-data-2/#:~:text=t%C3%A0Am%20produzido%20uma%20massiva%20quantidade,na%20Terra%20a%20cada%20segundo/>>. Acesso em: 15 jun. 2020.

GONÇALVES, MARCIO. Extração de Dados para Data Warehouse. Rio de Janeiro: Axcel Books, 2003.

GUERRA, Gabriela Netto. A IMPORTÂNCIA DO DATA WAREHOUSE NO PROCESSO DE TOMADA DE DECISÃO. 2005. 58 f. TCC (Graduação) - Curso de Sistemas de Informação, Fundação Getúlio Vargas – Fgv Management, Brasília – DF, 2005

INMON, W. H.; WELCH, J. D.; GLASSEY, K. L. Gerenciando data warehouse. São Paulo: Makron Books, 1999

INMON, William H. Como construir o data warehouse. Campus, 1997.

KIMBALL, R; ROSS, M.; THORNTHWAITE, W. The data warehouse lifecycle toolkit. 2. ed. Indianapolis, Indiana: Wiley Publishing Inc., 2008.

MACHADO, F. N. R. Tecnologia e Projeto de Data Warehouse: Uma visão multidimensional. Tatuapé: Érica, 2006.

MACHADO, Felipe Nery Rodrigues. Tecnologia e Projeto de Data Warehouse: Uma visão multidimensional. São Paulo: Editora Érica Ltda, 2004.

MAZZOLA, Irany Salgado. 2002. 153p. Projeto de Data Warehouse Dimensional. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Santa Catarina, Florianópolis, 2002. Disponível em: <<https://repositorio.ufsc.br/bitstream/handle/123456789/83465/184713.pdf?sequence=1/>>. Acessado em: 3 set. 2016.

MUSICANTE, Martin A. (ed). Data warehousing e OLAP. Curitiba: UFPR, 2001. 7p

MUSICANTE, Martin A. (ed). Data warehousing e OLAP. Curitiba: UFPR, 2001. 161p.

ORTEGA, Flávia. Entenda a atuação extrajudicial do Ministério Público. Disponível em: <<https://draflaviaortega.jusbrasil.com.br/noticias/322982018/entenda-a-atuacao-extrajudicial-do-ministerio-publico/>>. Acesso em: 15 jun. 2020.

PENTAHO, Data Integration - Kettle. Disponível em: <<http://kettle.pentaho.com/>> Acesso em: 28 de abr. 2020.

POSTGRESQL, PostgreSQL: The World's Most Advanced Open Source Relational Database. Disponível em: <<https://www.postgresql.org/>> Acesso em: 28 de abr. 2020.

PRIMAK, (2008). Decisões com BI. Disponível no e-book em: http://books.google.com.br/books/about/Decis%C3%B5es_com_BI_Business_Intelligence.html?id=cGBneHPeLvkC

SASSI, R. J. Análise das Modalidades de um Data Warehouse. Sinergia, Revista do Centro Federal de Educação Tecnológica de São Paulo, vol. 4, n. 1, 2003.

SHAMS K., Farishta M. (2001), “Data Warehousing: Toward knowledge Management”, Topics in Health Information Management, v. 21, n. 3, p. 24-32

VILARINO, Rafael Oliveira. Extração de Informações Estratégicas da Base de Dados do ENADE 2014. 2016. 33 f. Trabalho de Conclusão de Curso (Graduação) – Curso de Sistemas de Informação, Centro Universitário Luterano de Palmas, Palmas/TO, 2016.