



UNISUL

UNIVERSIDADE DO SUL DE SANTA CATARINA

CLÁUDIO REIS

**UMA ARQUITETURA DE BI: PORTAL DE ACESSO À INFORMAÇÃO E
TRANSPARÊNCIA DO TJSC – GESTÃO DE PESSOAS**

Florianópolis

2018



UNIVERSIDADE DO SUL DE SANTA CATARINA
CLÁUDIO REIS

**UMA ARQUITETURA DE BI: PORTAL DE ACESSO À INFORMAÇÃO E
TRANSPARÊNCIA DO TJSC – GESTÃO DE PESSOAS**

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: Prof. Aran Bey Tcholakian Morales, Dr Eng.

Florianópolis

2018

CLÁUDIO REIS

**UMA ARQUITETURA DE BI: PORTAL DE ACESSO À INFORMAÇÃO E
TRANSPARÊNCIA DO TJSC – GESTÃO DE PESSOAS**

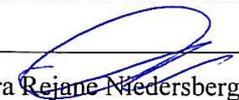
Este trabalho de conclusão de curso foi julgado adequado à obtenção do título de Bacharel em Sistemas de Informação e aprovado em sua forma final pelo Curso de Graduação em Sistemas de Informação da Universidade do Sul de Santa Catarina.

Florianópolis, 21 de novembro de 2018.



Professor Aran Bey Tcholakian Morales, Dr.

Universidade do Sul de Santa Catarina



Professora Vera Rejane Niedersberg Schuhmacher, Dra.

Universidade do Sul de Santa Catarina



Professora Daniella Vieira, MEng.

Universidade do Sul de Santa Catarina

Dedico este projeto aos usuários com os quais tive o prazer de trabalhar.

AGRADECIMENTOS

Agradeço à Dona Elsa da Silva, minha mãe, que sempre me mostrou o valor do estudo.

Agradeço aos mestres, que com paciência e sapiência, guiaram este incauto discípulo pelos caminhos do conhecimento.

Agradeço em especial à professora Vera Schuhmacher, à professora Maria Inés Castiñeira e ao mestre Aran Morales pela motivação que provocaram em mim. Sem eles eu não teria conseguido!

Agradeço à Maria do Carmo Olívia Provesano pelo o apoio nesta caminhada rumo à graduação.

Agradeço aos meus filhos, Rian Provesano Reis e Raíssa Provesano Reis que além de parceiros na vida, se tornaram parceiros nos empreendimentos também, inclusive neste projeto.

Agradeço aos usuários que fizeram parte da minha história profissional, com os quais aprendi muito, principalmente sobre pessoas.

Agradeço ao meu grupo de amigos, chamado de Turma das Quinta (nos encontramos todas às quintas há 32 anos) pela parceria. Aprendi muito com eles.

Agradeço aos companheiros de CPD do Tribunal de Justiça que sempre me inspiraram a seguir na área de tecnologia. Em especial à Lizete Helena Vieira, que me ensinou meu primeiro comando em Mumps e ao João Luiz Moreira, que foi fonte de inspiração para que eu seguisse na profissão de analista de sistemas.

Agradeço ao Júlio Cesar da Rocha de Castro, gerente na Unimed Florianópolis, durante o período que supervisionei o CPD. Um grande chefe que se tornou um grande amigo e acreditou no meu potencial para ser analista de sistemas num tempo em que a profissão não era tão conhecida.

Um agradecimento especial à empresa que criei em 2000, a Conceitual Sistemas de Informação. Nela aprendi que além de um excelente profissional, é necessário ser uma grande pessoa e tive a oportunidade de trabalhar com a Ana Cristina Pereira, o Márcio Soares, o Patrick Machado, a Priscilla Padilha, entre outros grandes profissionais.

Por fim, agradeço a todos que de uma maneira ou de outra, participaram do meu crescimento profissional e pessoal.

Muito Obrigado!

RESUMO

A transparência e o acesso à informação nunca estiveram tão em evidência em nossa sociedade. Prova disso é um sem-número de observatórios sociais (organizações da Sociedade Civil que tem por objetivo a fiscalização de órgãos públicos como prefeituras, por exemplo) existentes.

No Brasil, os órgãos públicos nas esferas Federal, Estadual e Municipal e no âmbito dos Três Poderes, devem disponibilizar seus dados para garantia da transparência e o acesso à informação à toda a população. No entanto, por serem assuntos relativamente novos, existe muita dificuldade em obter informação, mesmo que o acesso a estes dados seja regulamentado por lei.

Na alçada do Judiciário, quem garante o acesso à informação e à transparência é o Conselho Nacional de Justiça mais conhecido como CNJ. É ele que por instrumentos legais, chamados de Resoluções, dá visibilidade ao trabalho e gastos deste importante Poder Constituído da República. “Com a Lei de Acesso, a publicidade tornou-se a regra e o sigilo, a exceção”, lê-se no site do CNJ.

A Resolução 219/2016 do CNJ trata sobre a disposição e ocupação de cargos nas unidades judiciárias que compõem os tribunais de justiça pelo país. Esta resolução permite à Sociedade Civil Catarinense, ter acesso aos dados que tratam da distribuição pelas comarcas catarinenses do quadro de pessoal do Tribunal de Justiça de Santa Catarina. Apesar da facilidade de acesso aos dados, há uma dificuldade de se obter informação.

Este projeto coletou dados nas tabelas definidas pela Resolução 219/2016 do CNJ com o objetivo de transforma-los em informação. Estes dados foram “encaixados” em um modelo dimensional, utilizando-se os conceitos de *Business Intelligence*, que permitisse a visualização e inferências de informações relevantes no Portal de Acesso à Informação e Transparência – Gestão de Pessoas do Tribunal de Justiça de Santa Catarina. A ferramenta Power BI da Microsoft foi utilizada como forma de apresentação destas informações.

Este projeto facilitou o acesso a informações que estavam “escondidas” dentro dos dados disponíveis nas TLPs, por meio de visualizações e inferências criadas, executando assim sua proposta inicial de transformar dados em informações que pudessem ser facilmente alcançadas e facilitar a transparência ao cidadão.

Palavras-chave: *Business Intelligence*, Acesso à Informação, Transparência, CNJ, Resolução 219/2016. TJSC, Modelo Dimensional, Sistema de Apoio à Decisão.

“Além da Tecnologia. As pessoas.” (Cláudio Reis, 2010).

LISTA DE ILUSTRAÇÕES

Figura 1 – Componentes de um Sistema de Informação	19
Figura 2 – Elementos Básicos do <i>data warehouse</i>	26
Figura 3 – Representação do <i>Star Schema</i> (Modelo Estrela).....	30
Figura 4 – Proposta da solução escolhida	41
Figura 5 – Dimensão TLP	46
Figura 6 – Dimensão Geografia	47
Figura 7 – Dimensão Tempo	48
Figura 8 – Tabela Fato Locação	49
Figura 9 – O Processo de desenvolvimento	55
Figura 10 – Acesso à Informação e Transparência	56
Figura 11 – Acesso à Informação e Transparência Opção dos Semestres	57
Figura 12 – Acesso à Informação e Transparência Opção Tabelas	58
Figura 13 – Acesso à Informação e Transparência Opção Tabela (TLP 1 em PDF)	58
Figura 14 – Aplicativo Mobile Adobe Acrobat Reader	59
Figura 15 – Aplicativo Acrobat Reader com a opção Exportar PDF	60
Figura 16 – Aplicativo Adobe Acrobat Reader, opção Exportar PDF (TLP 1)	60
Figura 17 – Exportação finalizada (Planilhas Editadas e Unificadas)	61
Figura 18 – Modelo Dimensional Lógico	62
Figura 19 – Desenho genérico de ETL	63
Figura 20 – Carga da tabela dimensão TLP	64
Figura 21 – Carga da tabela fato	64
Figura 22 – Opção Obter Dados no Power BI da Microsoft	65
Figura 23 – Campo Munic (município) categorizado para Cidade	66
Figura 24 – Primeira inferência, Municípios com ArcGIS	67
Figura 25 – Segunda inferência, cruzamento: Municípios e CC	68
Figura 26 – Terceira inferência, filtro para municípios com mais de 20 CC	68
Figura 27 – Quarta inferência, cruzamento entre CC e LP (Lotação Paradigma) > zero	69
Figura 28 – Quinta inferência, filtro para LP acima de 60	69
Figura 29 – Sexta inferência, gráfico barras, Ano, Primeiro Semestre e número de CC	70
Figura 30 – Sétima inferência, gráfico de colunas, Ano, Primeiro Semestre somatório	70
Figura 31 – Oitava inferência, colunas agrupadas, Ano, Primeiro Semestre somatório	71

LISTA DE TABELAS

Tabela 1 – Descrição dos atributos dos requisitos funcionais	44
Tabela 2 – Descrição dos atributos dos requisitos não-funcionais	45
Tabela 3 – Descrição dos atributos da tabela dimensão TLP	46
Tabela 4 – Descrição dos atributos da tabela dimensão Geografia	47
Tabela 5 – Descrição dos atributos da tabela dimensão Tempo	48
Tabela 6 – Descrição dos atributos da tabela fato locação	49

LISTA DE SIGLAS

ArcGIS – *Geographic Information System*

BD – Banco de Dados

BI – *Business Intelligence*

CC – Cargo Comissionado

CNJ – Conselho Nacional de Justiça

CPD – Centro de Processamento de Dados

DBMS – *Database Management System*

DM – *Data Mart*

LP – Lotação Paradigma

ETL – *Extraction, Transformation and Load*

E-R – Entidade-Relacionamento

FC – Função em Confiança

FK – *Foreign Key*

JDBC – *Java Database Connectivity*

ODS – *Operational Data Store*

OLAP – *On-Line Analytic Processing*

PDF – *Portable Document Format*

PK – *Primary Key*

SGBD – Sistema de Gerenciamento de Banco de Dados

SI – Sistema de Informação

SQL – *Structured Query Language*

TJSC – Tribunal de Justiça de Santa Catarina

TLP – Tabela de Lotação de Pessoal

XLS – *Extensible Style Language*

3NF – Terceira Forma Normal

SUMÁRIO

1 INTRODUÇÃO	13
1.1 Problemática	14
1.2 Objetivos	15
1.2.1 Objetivo Geral	15
1.2.2 Objetivos Específicos	15
1.3 Justificativa	16
1.4 Estrutura do Projeto	17
2 REFERÊNCIA BIBLIOGRÁFICA	18
2.1 Conceitos de Sistemas de Informação	18
2.2 Conceitos de <i>Business Intelligence</i>	20
2.2.1 A Importância de um Projeto de <i>Business Intelligence</i>	21
2.2.2 ETL - Extração, Transformação e Carga.....	21
2.2.3 <i>Data Warehouse</i>	23
2.2.3.1 Requisitos para um <i>Data Warehouse</i>	25
2.2.3.2 Componentes de um <i>Data Warehouse</i>	26
2.2.3.3 Granularidade	27
2.2.4 Modelagem Dimensional	27
2.2.4.1 Modelo Estrela, Fatos, Dimensões, Medidas e Agregados.....	29
2.2.5 OLAP – Aplicações Analíticas.....	36
2.2.5.1 Ferramentas OLAP.....	36
2.3. Considerações finais sobre o capítulo	37
3 METODOLOGIA	39
3.1 Tipo de Pesquisa	39
3.2 Etapas Metodológicas	40
3.3 Proposta da Solução.....	41
3.4 Delimitação do Projeto.....	42

4 MODELAGEM.....	43
4.1 Requisitos.....	43
4.1.1 Requisitos Funcionais.....	43
4.1.2 Requisitos Não-Funcionais.....	45
4.2 Modelagem Dimensional.....	45
4.2.1 Dimensão TLP.....	46
4.2.2 Dimensão Geografia.....	47
4.2.3 Dimensão Tempo.....	48
4.2.4 Tabela Fato Locação	48
4.3 Considerações finais sobre o capítulo.....	50
5 DESENVOLVIMENTO.....	51
5.1 Tecnologias & Ferramentas.....	51
5.1.1 Google Chrome	51
5.1.2 Adobe Acrobat Reader	51
5.1.3 Microsoft Excel	52
5.1.4 SQL Power Architect	53
5.1.5 PostgreSQL.....	53
5.1.6 Pentaho Data Integration (Kettle)	53
5.1.7 Power BI	54
5.2 Arquitetura do BI	54
5.2.1 O Processo	55
5.2.2 Download das tabelas em PDF.....	56
5.2.3 Conversão das tabelas em PDF para Excel	59
5.2.4 Edição das planilhas em Excel	61
5.2.5 O modelo dimensional	61
5.2.6 O ETL	63
5.2.6.1 A carga	63
5.2.7 A visualização e inferências	65
5.3 Considerações finais sobre o capítulo	71

6 CONCLUSÕES E TRABALHOS FUTUROS.....	72
6.1 Conclusões	72
6.2 Trabalhos futuros	72
REFERÊNCIAS	74
ANEXOS	75
ANEXO A	76

1 INTRODUÇÃO

Todo cidadão tem o direito constitucional de obter dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, conforme prevê o artigo 5º, inciso XXXIII da Constituição Federal. Para isso foi criada a lei que regulamenta o acesso à informação, conhecida como Lei de Acesso à Informação. Para garantir o cumprimento da lei pelo Judiciário, o Conselho Nacional de Justiça publicou a Resolução 215/2015 que determina a utilização de procedimentos objetivos e ágeis, de forma transparente, clara em linguagem de fácil compreensão, a fim de garantir às pessoas naturais e jurídicas o direito de acesso à informação (CNJ, 2018).

O Conselho Nacional de Justiça e as unidades judiciárias do Brasil dão transparência a seus processos internos e ações por meio de leis, resoluções e recomendações. O objetivo dessa política de visibilidade é deixar acessíveis aos cidadãos todas as informações sobre o trabalho do órgão e os gastos da Justiça brasileira (CNJ, 2018).

A Resolução 219/2016 do CNJ dispõe sobre a distribuição de servidores, de cargos em comissão e de funções de confiança nos órgãos do Poder Judiciário de primeiro e segundo graus (CNJ, 2018).

No Portal de Acesso à Informação e Transparência do Tribunal de Justiça de Santa Catarina - Tabela de Lotação de Pessoal é possível, à Sociedade Catarinense, acompanhar a distribuição dos recursos humanos (servidores). Lá, em abas que representam o período de seis meses anteriores, encontramos as tabelas que referenciam as unidades de apoio aos serviços jurisdicionais executados pelo Tribunal de Justiça de Santa Catarina, nos quais os servidores estão alocados. Estas unidades são Judiciárias de Primeiro e Segundo Grau (TLP 1), de Apoio Direto (TLP 2), e de Apoio Indireto (TLP 3) às Atividades Judicantes. As tabelas estão organizadas de acordo com as orientações e definições da Resolução 219/2016 do Conselho Nacional de Justiça. Dentro das tabelas é possível encontrar o nome da própria tabela, um mini “glossário” dos dados tabelados naqueles documentos, além dos próprios dados. Os documentos estão no formato *Portable Document Format*, palavras em inglês para Formato Portátil de Documento, sigla PDF.

1.1 PROBLEMÁTICA

Apesar de fácil acesso às tabelas reguladas pela Resolução 219/2016 no Portal de Acesso à Informação e Transparência do Tribunal de Justiça de Santa Catarina, existe uma grande dificuldade na visualização e na inferência de informações dos dados que trata a resolução do Conselho Nacional de Justiça e que constam no Portal. Estando as tabelas no formato PDF, não é possível a realização de nenhuma operação direta sobre elas, como por exemplo: soma, média, divisão e etc. Ainda que possível fosse a realização destas operações, os dados que lá se encontram, estão na forma analítica de visualização, dificultando o encontro de informações relevantes. Existe ainda outro agravante, as ocorrências de grau, tipo de unidade e município, que servem claramente para agregar diversas ocorrências, estão em mais de uma linha, dificultando mais uma vez, a visualização e inferência de informações, e levando a conclusões incompletas ou equivocadas. Dados como grau, tipo de unidade e município, além da visualização e inferência de ocorrência unívoca, devem permitir sumarização para fácil apuração, o que não ocorre na forma em que são apresentadas as tabelas TLP 1 – Unidades Judiciárias de Primeiro e Segundo Graus, TLP 2 – Demais Unidades de Apoio Direto à Atividade Judicante e Unidades de Apoio Indireto à Atividade Judicante. Lembrando que transparência é o que não é ambíguo, traz clareza e limpidez.

A construção de uma arquitetura de BI que extraia as informações necessárias das tabelas TLP 1, TLP 2 e TLP 3, pode ajudar na transparência e na confiança das informações prestadas pelo Portal.

Segundo Barbieri (2001), o conceito de BI deve ser entendido como o processo de desenvolvimento de estruturas especiais de armazenamento de informações como *Data Warehouse* (DW), *Data Marts* (DM) e ODS (*Operational Data Store*) com o objetivo de se montar uma base de recursos informacionais, capaz de sustentar a camada de inteligência da empresa e possível de ser aplicada aos seus negócios, como elementos diferenciais e competitivos. Podemos claramente, aplicar este conceito às tabelas disponível no site do Tribunal de Justiça de Santa Catarina.

1.2 OBJETIVOS

Com base na apresentação do problema são apresentados os objetivos geral e específicos do presente projeto.

1.2.1 Objetivo Geral

Auxiliar através da construção de uma arquitetura de *BI-Business Intelligence*, o acesso de informações sintetizadas das tabelas (TLP 1 – Unidades Judiciárias de Primeiro e Segundo Graus, TLP 2 – Demais Unidades de Apoio Direto à Atividade Judicante e Unidades de Apoio Indireto à Atividade Judicante) da Resolução 219/2016 do CNJ que constam no Portal de Acesso à Informação e Transparência – Tabela de Lotação de Pessoal do Tribunal de Justiça de Santa Catarina.

1.2.2 Objetivos específicos

- a) Construir um modelo dimensional para as tabelas (TLP 1 – Unidades Judiciárias de Primeiro e Segundo Graus, TLP 2 – Demais Unidades de Apoio Direto à Atividade Judicante e Unidades de Apoio Indireto à Atividade Judicante) da Resolução 219/2016 do CNJ que constam no Portal de Acesso à Informação e Transparência – Tabela de Lotação de Pessoal do Tribunal de Justiça de Santa Catarina;
- b) Realizar o processo de ETL, Extração, Transformação e Carga dos dados das tabelas (TLP 1 – Unidades Judiciárias de Primeiro e Segundo Graus, TLP 2 – Demais Unidades de Apoio Direto à Atividade Judicante e Unidades de Apoio Indireto à Atividade Judicante) da Resolução 219/2016 do CNJ que constam no Portal de Acesso à Informação e Transparência – Tabela de Lotação de Pessoal do Tribunal de Justiça de Santa Catarina;

- c) Construir uma área de apresentação utilizando uma ferramenta apropriada para a visualização e inferências a serem realizadas nas tabelas (TLP 1 – Unidades Judiciárias de Primeiro e Segundo Graus, TLP 2 – Demais Unidades de Apoio Direto à Atividade Judicante e Unidades de Apoio Indireto à Atividade Judicante) da Resolução 219/2016 do CNJ que constam no Portal de Acesso à Informação e Transparência – Tabela de Lotação de Pessoal do Tribunal de Justiça de Santa Catarina.

1.3 JUSTIFICATIVA

Percebe-se que mesmo com os dados disponíveis no Portal de Acesso à Informação – Gestão de Pessoas do Tribunal de Justiça de Santa Catarina, a visualização e a inferência que resultem em informações relevantes aos cidadãos catarinenses estão em um nível de difícil extração. É um verdadeiro paradoxo: acesso fácil aos dados, dificuldade de acesso às informações, mesmo que a Lei esteja sendo cumprida.

Com isso podemos dizer que utilizar uma arquitetura de *Business Intelligence* para possibilitar o acesso a informações relevantes, parece ser o caminho mais natural possível, pois o BI é a transformação destes dados brutos em informações trabalhadas, e isto facilitará o acesso ao que realmente importa.

O tema apresentado é relevante por tratar de informações que devem estar disponíveis às pessoas para que eles possam exercer seu direito à cidadania, informando-se de como estão sendo utilizados os recursos (no caso do projeto, os recursos humanos) pelo Tribunal de Justiça de Santa Catarina.

Na abordagem deste tema, vislumbra-se o melhoramento no acesso e na qualidade das informações disponíveis no Portal de Acesso à Informação Gestão de Pessoas tabelas TLP 1, TLP 2 e TLP 3 no site do TJSC.

As vantagens e benefícios do tema abordado são: facilidade e transparência nas informações.

1.4 ESTRUTURA DO PROJETO

A estrutura deste projeto é de 6 capítulos a saber:

O primeiro capítulo apresentará a introdução, está subdividido em: introdução propriamente dita, apresentação do problema e objetivos (geral e específicos).

O segundo capítulo tratará dos conceitos necessários para a construção de Sistemas de Informação no processo decisório, abrangendo os conceitos de BI, *Data Warehouse*, modelo dimensional, extração, limpeza e carga dos dados e a ferramenta de visualização das informações.

O terceiro capítulo mostrará a metodologia da pesquisa, solução proposta e delimitações.

O quarto capítulo exporá a modelagem da solução proposta com seus requisitos funcionais e não-funcionais e modelo dimensional das tabelas (TLP 1 – Unidades Judiciárias de Primeiro e Segundo Graus, TLP 2 – Demais Unidades de Apoio Direto à Atividade Judicante e Unidades de Apoio Indireto à Atividade Judicante) da Resolução 219/2016 do CNJ que constam no Portal de Acesso à Informação e Transparência – Tabela de Lotação de Pessoal do Tribunal de Justiça de Santa Catarina.

O quinto capítulo tratará das tecnologias e ferramentas utilizadas no desenvolvimento deste projeto.

O sexto e último capítulo apresentará as conclusões e trabalhos futuros.

2 REFERÊNCIA BIBLIOGRÁFICA

Neste capítulo são abordados, os conceitos das tecnologias de informação no processo decisório, abrangendo *Business Intelligence* (BI) e os elementos que formam a arquitetura de uma solução de BI: o sistema ETL (extração, transformação e carga dos dados), o repositório de dados (*Data Warehouse* e o modelo dimensional) e as ferramentas de apresentação.

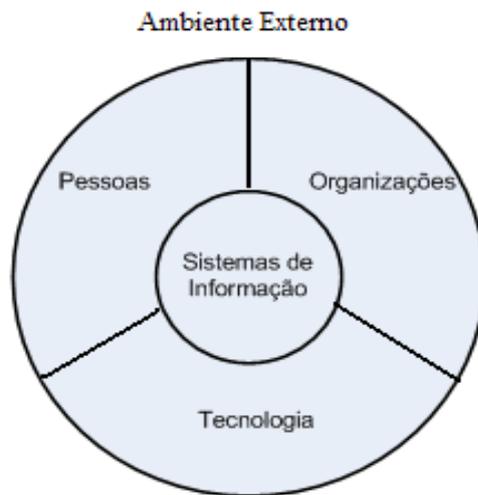
As decisões padecem de um mal inerente, o tempo. Para se tomar uma decisão que tenha um efeito positivo, deve-se fazê-lo no tempo adequado. Isto já é complicado na vida particular, quando lidamos com um número restrito de informações. Na vida empresarial, então o problema cresce exponencialmente, devido ao grande número de variáveis.

Durante muito tempo, os CPDs das empresas, procuraram fornecer as mesmas, o básico do básico na informática. Agilidade no processamento dos dados, gerando faturas, emitindo carteiras, calculando a folha de pagamento e assim por diante. Já se praticava nesta época o conceito de sistemas de informação. Mas o volume dos dados foi crescendo, e aí então, não bastava agilizar processos operacionais, mas dar suporte aos processos gerenciais.

2.1 CONCEITO DE SISTEMAS DE INFORMAÇÃO

Sistema de informação pode ser definido como um conjunto de componentes inter-relacionados trabalhando juntos para coletar, recuperar, processar, armazenar e distribuir informação com a finalidade de facilitar o planejamento, o controle, a coordenação, a análise e o processo decisório em empresas e outras organizações. (Laudon e Laudon, 1999, página 4).

Figura 1 : Componentes de um Sistema de Informação



Fonte: Laudon e Laudon (1999, página 4)

Sistemas de informação é o nome que se dá a um conjunto de componentes inter-relacionados, informáticos ou não, que melhorem a qualidade dos processos empresariais visando aumentar a competitividade das empresas.

Os sistemas de informação podem ser classificados em:

- **Sistemas de Informação Operacional:** tratam das transações rotineiras da organização; Comumente encontrados em todas as empresas automatizadas;
- **Sistemas de Informação Gerencial:** agrupam e sintetizam os dados das operações da organização para facilitar a tomada de decisão pelos gestores da organização;
- **Sistemas de Informação Estratégicos:** integram e sintetizam dados de fontes internas e externas à organização, utilizando ferramentas de análise e comparação complexas, simulação e outras facilidades para a tomada de decisão da cúpula estratégica da organização. Estes também conhecidos como Sistemas de Apoio à Decisão.

O papel fundamental dos sistemas de informação é dar suporte às operações, à tomada de decisão e à vantagem estratégica aos negócios, apoiando às operações, tomada de decisão e a vantagem estratégica.

2.2 CONCEITO DE *BUSINESS INTELLIGENCE*

A expressão *Business Intelligence*, que quer dizer inteligência nos negócios, foi criada pelo Gartner Group em 1992. *Business Intelligence* veio como uma evolução dos sistemas de apoio à decisão.

Para Barbieri (2001, página 34), o conceito de *Business Intelligence*, de forma mais ampla, pode ser entendido como a utilização das variadas fontes de informação para se definir estratégias de competitividade nos negócios da empresa.

Os dados estão amontoados nos mais diversos sistemas (informatizados ou não) nas empresas. Por serem muitos e estarem dispersos dificultam a sua consolidação e abstração para uma avaliação que suporte o processo decisório.

O objetivo maior das técnicas de *Business Intelligence* está exatamente na definição de regras e ferramentas para a formatação adequada destes volumes de dados, visando transformá-los em depósitos estruturados de informações, independentemente de sua origem, Barbieri (2001, página 34).

Para Rosini & Palmisano (2003, página 56), o *Business Intelligence* é uma nova ferramenta da tecnologia da informação, em que a maioria das corporações, ainda não está preparada para as implementações, pois são necessários os seguintes itens:

- Há necessidade de mudança e/ou adaptação da cultura da organização;
- É necessário o apoio, o incentivo e a cobrança do alto escalão da organização;
- É preciso que os indivíduos estejam capacitados tecnicamente (mediante treinamentos), não só no uso da ferramenta tecnológica, mas na sua própria formação individual e profissional.

A tecnologia da informação tem sido até agora uma produtora de dados, não de informação, e muito menos uma produtora de novas e diferentes questões e estratégias, em que os altos executivos não têm usados a nova tecnologia, porque ela não tem oferecido as informações de que eles precisam para suas próprias tarefas. Outro grande problema enfrentado pelos executivos é o de ter todos os dados espalhados por inúmeros sistemas, não havendo integração entre a maioria deles. Não podemos esquecer que as decisões estratégicas devem ser tomadas em um período curto de tempo, pois o mercado movimenta-se rapidamente, e um dia ou mais, pode representar grandes prejuízos para uma empresa.

Ainda, para Rosini e Palmisano (2003, página 56), o *business intelligence* não deixa de ser um processo de coleta, transformação, análise e distribuição de dados, para melhorar a decisão dos negócios pertinentes às organizações, onde os bancos de dados são a infra-estrutura básica de qualquer sistema de *bussiness intelligence*, pois é onde estarão armazenados os dados que serão transformados em informações relevantes.

2.2.1 A Importância de um projeto de *Business Intelligence*

Transformar numerosos dados, que estão espalhados pela organização, em informação aplicável aos negócios da empresa, para que ela tenha alguma vantagem competitiva ou aumento de produtividade na entrega de um serviço (quando se trata de empresa pública), é uma tarefa árdua. E o resultado desta tarefa nem sempre é confiável. Com um projeto de *business intelligence*, o trabalho de consolidação dos dados é drasticamente reduzido e a confiabilidade neles aumenta bastante.

Outro ponto importante com a implantação de um projeto de *business intelligence* diz respeito a velocidade de respostas no auxílio do processo decisório. Na maioria das vezes estas respostas devem ser conseguidas quase que instantaneamente ante os questionamentos que surgem na vida cotidiana das empresas.

2.2.2 ETL - Processo de Extração, Transformação e Carga

Segundo Kimball (2002, página 10), o ETL (*Extract-Transformation-Load*), é um conjunto de processos, que ocorre em uma área de armazenamento do *data warehouse*, conhecida como *Data Staging Area*.

O principal requisito de arquitetura da *data staging area* é que ela não esteja acessível aos usuários e que não forneça serviços de consulta nem de apresentação. (Kimball 2002, página 10).

O “E” (de extração) da sigla ETL Kimball (2002, página 10) define como sendo “ a primeira etapa do processo de obtenção de dados no ambiente de *data warehouse*. O processo de extração envolve a leitura e a compreensão de dados de origem e cópia dos dados necessários ao *data warehouse* na *staging area* para que sejam manipulados posteriormente”.

Com relação a extração, Inmon (1997, página 74) alerta que “A idéia de que a construção de um *data warehouse* consiste meramente em um processo de extração tem como erro básico o fato de que os dados existentes no ambiente operacional não são integrados.”, e ainda continua “[...] tentar extrair dados dos diversos lugares em que eles existem é um problema muito complexo”.

Após a Extração, é a vez do “T” de transformação. Os dados após extraídos de diversas fontes, podem apresentar variações, como por exemplo: em uma fonte o gênero pode ser “M” para masculino e “F” para feminino, e em outra fonte poder ser “H” para homem e “M” para mulher. Apesar das duas fontes estarem tratando de gênero (comumente chamado de sexo), cada fonte representa o dado de uma maneira diferente. A transformação se encarrega de padronizar esta informação. Em Inmon (1997, página 74), para o exemplo acima, ele cita “[...] não importa como será feita a codificação do gênero no *data warehouse*, desde que ela seja feita de forma coerente. À medida que os dados passam para o *data warehouse*, os diferentes valores precisam ser corretamente decodificados e recodificados com o valor apropriado.

“Depois que os dados são extraídos para a *staging area*, ocorrem muitas transformações em potencial, como filtragem dos dados (correções de erros de digitação, solução de conflitos de domínio, tratamento de elementos ausentes ou a divisão em formatos padrão), combinação de dados das várias origens, cancelamento de dados duplicados e atribuição de chaves de *warehouse*. Essas transformações são todas precursoras para carregar os dados na área de apresentação do *data warehouse*.” Kimball, (2002, página 10)

A última etapa do ETL, é o processo de Carga (*load* em inglês, por isso “L”). Este processo irá alimentar o *data warehouse*, depois dos dados extraídos e transformados. A importância desta etapa está no fato de que ela servirá para criar a camada de apresentação do *data warehouse*. Segundo Kimball (2002, página 12), “O processo de carga no ambiente do *data warehouse* normalmente assume a forma de um processo de apresentação de tabelas dimensionais garantidas por controle de qualidade para os vários recursos de carga [...]”

Para Inmon (1997, página 76), há três tipos de carga que podem ser feitos do ambiente operacional para o *data warehouse*:

- O carregamento de dados históricos;

- O carregamento de dados de valor corrente no ambiente operacional;
- O carregamento de alterações do *data warehouse* a partir de alterações (atualizações) que tenham ocorrido no ambiente operacional desde a última atualização do *data warehouse*.

Ainda, para Inmon (1997, página 76), como regra, o carregamento de dados históricos representa um desafio menor uma vez que ele não é feito com frequência.

Da mesma forma, o carregamento de dados do ambiente operacional existente também não constitui grande desafio porque precisa ser feito apenas uma vez. Em geral, o ambiente de sistemas operacionais existentes pode ser descarregado em um arquivo sequencial e este pode ser descarregado no *data warehouse* sem acarretar danos ao ambiente *on-line*.

O Carregamento de dados durante o processamento normal – enquanto são efetuadas alterações sobre o ambiente operacional – consiste no maior desafio ao arquiteto de dados. Não é fácil realizar o rastreamento eficiente e o tratamento dessas alterações. A varredura de arquivos existentes é, portanto, uma importante questão a ser enfrentada pelo arquiteto do *data warehouse*.

Ferramentas de ETL podem ser compradas (há opções no mercado) ou podem ser desenvolvidas dentro do projeto de *data warehouse*, nesta opção há necessidade de um grande conhecimento dos dados operacionais.

A etapa de ETL deve ser bem conduzida, pois como se trata de uma parte crítica do projeto de *data warehouse*, ele pode comprometer o resultado do mesmo.

2.2.3 DATA WAREHOUSE

Para começar a definir o que é um *data warehouse*, serão colocados alguns questionamentos: como escolher uma estratégia de ação, convicto que terei resultados positivos? Como conhecer o melhor custo/benefício para uma ação a ser tomada? Tenho, como visualizar as informações por assunto de forma temporal? O que o tempo pode me dizer a respeito do desempenho da minha empresa?

Um *data warehouse* é isso, um banco de dados capaz de dar estas e outras respostas aos questionamentos de gerentes e diretores das empresas. Ele é uma poderosa ferramenta de *business intelligence* que pode orientar o caminho para o aumento de produtividade na busca por melhores resultados.

“Um dos bem mais preciosos de qualquer empresa são suas informações. E quase sempre a empresa mantém esse bem de duas formas: os sistemas operacionais de registro e o *data warehouse*”. Kimball (2002, página 2)

Segundo Inmon (1997, página 33) “Um *data warehouse* é um conjunto de dados baseado em assuntos, integrado, não-volátil, e variável em relação ao tempo, de apoio às decisões gerenciais.”

“*Data warehouse*, cuja a tradução literal é Armazém de Dados, pode ser definido como um banco de dados, destinado a sistemas de apoio à decisão e cujos dados foram armazenados em estruturas lógicas dimensionais, possibilitando o seu processamento analítico por ferramentas especiais (OLAP e *Mining*)”. Barbieri (2001, página 49)

O objetivo de um *data warehouse* é obter um grande depósito de informações para utilização em aplicações não transacionais, por isso o denominamos de sistema de apoio à decisão, o qual deve apresentar informações de séries históricas, que refletem a evolução de fatos do dia-a-dia de negócios de uma organização. Machado (2004, página 97)

Inmon (1997, página 33), explica as características de um *data warehouse*:

- Baseado em assuntos. Os sistemas operacionais clássicos são organizados em torno das aplicações da empresa. No caso de uma companhia de seguros, as aplicações podem ser automóvel, saúde, vida e perdas. Os principais assuntos ou negócios da empresa podem ser cliente, apólice, prêmio e indenização.
- Integrado. A integração ocorre quando os dados passam do ambiente operacional baseado em aplicações para o *data warehouse*. [...] O processo de introdução dos dados no *data warehouse* é conduzido de forma que as muitas inconsistências das aplicações sejam desfeitas. [...]
- Não-volátil. [...] No ambiente operacional, os dados sofrem atualizações. Contudo, os dados existentes no *data warehouse* apresentam um conjunto de características muito diferentes. Os dados no *data warehouse* são carregados (normalmente em grandes quantidades) e acessados, Mas a atualização dos dados (geralmente) não ocorre no ambiente do *data warehouse*.

- Variável em relação ao tempo. O horizonte de tempo válido para o *data warehouse* é significativamente maior do que o dos sistemas operacionais. Um horizonte de tempo de 60 a 90 dias é o normal para os sistemas operacionais; um horizonte de tempo de 5 a 10 anos é o normal para o *data warehouse*; Bancos de dados operacionais contêm dados de “valor corrente” – dados cuja exatidão é válida para o momento de acesso. Assim sendo, dados de valor corrente podem ser uma série sofisticada de instantâneos, capturados num determinado momento; A estrutura de chave dos dados operacionais pode conter, ou não, elementos de tempo, como ano, mês, dia etc. A estrutura de chave do *data warehouse* sempre contém algum elemento de tempo.

2.2.3.1 REQUISITOS PARA UM *DATA WAREHOUSE*

Para Kimball (2002, página 3), os requisitos de um *data warehouse* são:

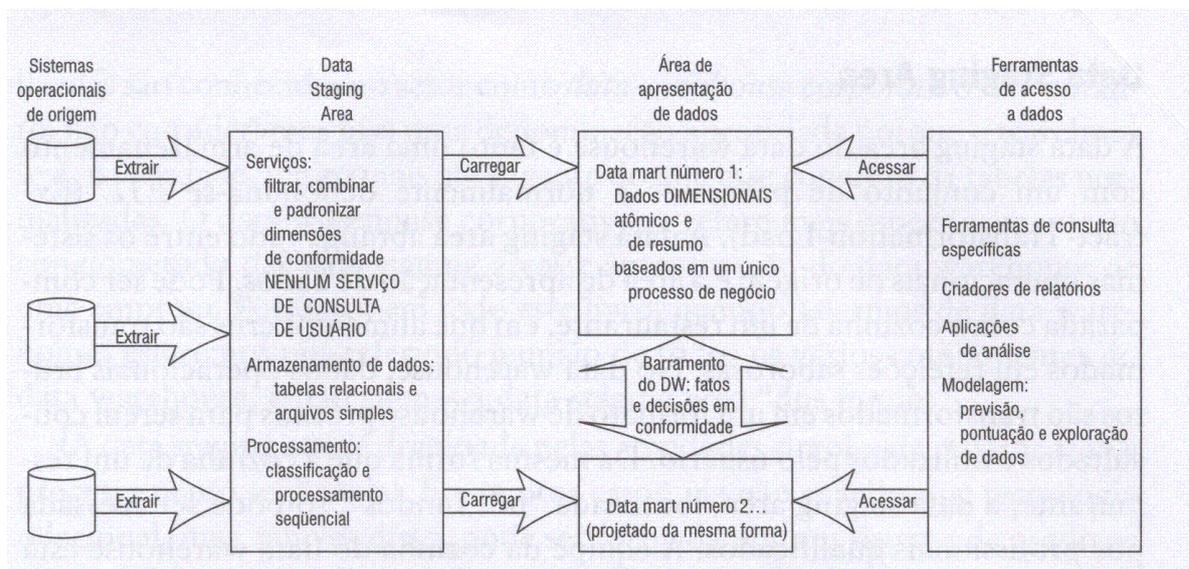
- **Fazer com que informações de uma empresa possam ser facilmente acessadas.** O conteúdo do *data warehouse* deve ser compreensível. Os dados devem ser intuitivos e óbvios para o usuário da área de negócios e não apenas para o desenvolvedor. [...] Os usuários das áreas de negócio querem separar e combinar os dados no *warehouse* em combinações infinitas, um processo que normalmente costuma ser denominado “separação e combinação” (*slicing and dicing*). [...]
- **Apresentar informações da empresa de modo consistente.** Os dados no *warehouse* devem ser confiáveis. Eles devem ser obtidos cuidadosamente a partir de várias fontes na empresa, filtrados, submetidos a um controle de qualidade e liberados apenas quando estiverem prontos para serem usados. [...]
- **Ser Adaptável e flexível a mudanças.** [...] As necessidades dos usuários, as condições comerciais, os dados e a tecnologia, todos estão sujeitos às mudanças decorrentes do tempo. [...]
- **Ser um baluarte seguro que protege as informações.** O tesouro de informações da empresa está guardado no *data warehouse*. [...]
- **Funcionar como base para uma melhor tomada de decisões.** O *data warehouse* deve conter os dados apropriados para dar suporte à tomada de decisões. [...]

- **Fazer com que a comunidade de negócio, aceite-o.** [...] Se a comunidade de negócio não adotar o *data warehouse* e continuar a usá-lo ativamente por seis meses após o treinamento, então isso indica que ele não passou no teste de aceitação. [...]

2.2.3.2 COMPONENTES DE UM *DATA WAREHOUSE*

Para Kimball (2002, página 8), vale a pena compreender bem cada um dos componentes do *data warehouse*, antes de começar a criá-lo. Cada componente atende a uma função específica, e é preciso aprender a importância estratégica de cada um deles e como controlá-lo de modo eficiente para obter êxito com o *data warehouse*. Confundir os papéis e as funções de cada componente, pode significar o fracasso do projeto.

Figura 2 : Elementos Básicos do *data warehouse*



Fonte: Kimball (2002, página 7)

Segundo Kimball (apud Machado 2004, página 81), desenvolver um *data warehouse* é uma questão de casar as necessidades dos seus usuários com a realidade dos dados disponíveis.

2.2.3.3 GRANULARIDADE

Granularidade de dados refere-se ao nível de sumarização dos elementos e de detalhe disponíveis nos dados, considerado o mais importante aspecto do projeto de um *Data Warehouse*. Machado (2004, página 59)

A granularidade trata da forma de agrupar informações. Ela pode ser baixa granularidade (alto nível de detalhe), ou ser alta granularidade (baixo nível de detalhe).

Segundo Machado (2004, página 60), a granularidade está diretamente ligada a capacidade de recuperar a informação em um *Data Warehouse* e por consequência atinge direto a performance do DW. Ele observa ainda que a granularidade está intimamente ligada a todos os fatores de classificação das informações que estiverem sendo utilizadas.

2.2.4 MODELAGEM DIMENSIONAL

Segundo Kimball (2002, página 13), “Modelagem dimensional é um novo nome para uma técnica antiga que permitia tornar os bancos de dados fáceis e compreensíveis. Caso após caso, a partir da década de 1970, as empresas de TI, as consultorias, os usuários finais e os fornecedores migraram para uma estrutura dimensional simples para atender à necessidade humana fundamental de simplicidade”.

Como exemplo, ele cita uma empresa que tivesse em sua atividade principal a venda de produtos em mais de um mercado (região) e já algum tempo. Poderíamos imaginar uma figura multidimensional (um cubo) com as seguintes faces: produto, região e tempo. “Os pontos dentro do cubo são o local em que as medições dessa combinação de produto, mercado e tempo está armazenada”.

Já para Barbieri (2001, página 74), a modelagem dimensional é a técnica de projeto que conduz os dados a formação de um modelo onde a informação reside na interseção de várias dimensões (pontos do cubo).

Os pontos dentro do cubo, são os locais em que as medições dessa combinação de produto, mercado e tempo acontecem. A capacidade de visualizar algo tão abstrato como um conjunto de dados de forma concreta e tangível é o segredo da capacidade de compreensão.

A modelagem dimensional é bem diferente da modelagem 3NF. A modelagem 3NF é uma técnica de design que tenta remover redundâncias de dados. Os dados são divididos em muitas entidades discretas, e cada uma delas se torna uma tabela em um banco de dados relacional.

Tanto o modelo 3NF (será chamado daqui para a frente de normalizado ou relacional) como o Dimensional, podem ser representado por diagramas de Entidade Relacionamento (diagramas ER), que a principal diferença entre os dois modelos está no nível de normalização dos modelos. E que a modelagem dimensional se aplica tanto a banco de dados relacionais bem como a banco de dados multidimensionais. Kimball (2002, página 13)

A modelagem dimensional permite que os usuários percebam os dados numa forma mais próxima de seu entendimento, com várias perspectivas possíveis, dentre elas o tempo e o espaço.

O produto final da modelagem Dimensional é um modelo conceitual dimensional, formado por tabelas Fato e tabelas Dimensão.

O projetista de *data warehouse*, deve criar os *data marts* com a melhor granularidade possível para que as consultas de usuários não fiquem comprometidas. Todos os *data marts* devem ser construídos usando dimensões e fatos diferentes, aos quais nos referimos como de conformidade. Essa é a base da arquitetura de barramento do *data warehouse* [...]. Sem dimensões e fatos compartilhados e em conformidade, um *data mart* é uma aplicação independente. Os *data marts* independentes que não podem ser reunidos representam a ruína do movimento do *data warehouse* porque simplesmente perpetuam visões incompatíveis da empresa.

A arquitetura de barramento, é o que dará consistência (objetivo maior) ao *data warehouse*. A arquitetura de barramento “é o método incremental e arquitetônico para criar o *warehouse* da empresa”. (Kimball 2002, página 15).

A ferramenta utilizada para criar, documentar e comunicar a arquitetura de barramento é a matriz de barramento.

Machado (2004, página 79), chama o modelo dimensional de modelo multidimensional, e o define assim “[...] a modelagem multidimensional, é mais simples, mais expressiva e mais fácil de entender que a modelagem ER [...]”.

Mais à frente ele explica que “A modelagem multidimensional é uma técnica de concepção e visualização de um modelo de dados de um conjunto de medidas que descrevem aspectos comuns de negócios. É utilizada especialmente para sumarizar e reestruturar dados e apresentá-los em visões que suportem a análise dos valores desses dados.”

Um modelo multidimensional é formado por três elementos básicos:

- Fatos;
- Dimensões;
- Medidas (variáveis);

2.2.4.1 MODELO ESTRELA, FATOS, DIMENSÕES, MEDIDAS E AGREGADOS

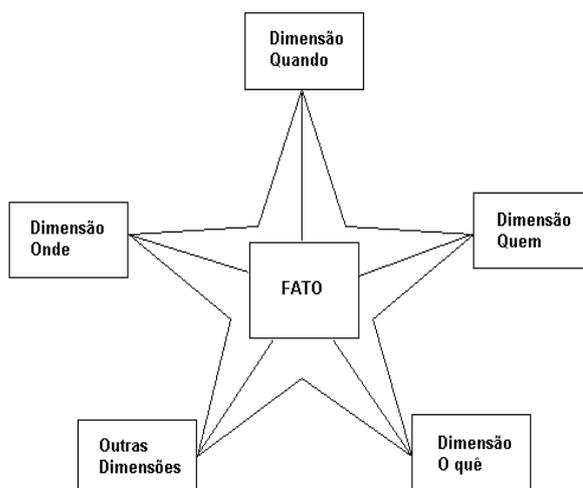
MODELO ESTRELA

O modelo estrela (*star schema*) é o termo para a designação de modelos de dados multidimensionais. Ele é a estrutura básica deste mesmo modelo de dados multidimensional.

Sua composição típica possui uma grande entidade central denominada fato (*fact table*) e um conjunto de entidades menores denominadas dimensões (*dimension tables*), arranjadas ao redor dessa entidade central, formando uma estrela.

Já Barbieri (2001, página 53), define o modelo estrela como sendo um conjunto de tabelas Fatos (onde se concentram os dados de interesse, passíveis de manipulação numérica e estatística), e tabelas Dimensão (tabelas satélites que possuem as chaves de entrada do modelo, além das informações descritivas de cada dimensão).

Figura 3: Modelo estrela (*star schema*)



Fonte: Machado (2004, página 93)

FATOS

No modelo dimensional, *Fato* representa uma medição de negócio. Kimball (2002, página 21) explica que “uma tabela de fatos é a principal tabela de um modelo dimensional em que as medições numéricas de desempenho da empresa estão armazenadas”. Deve haver um grande esforço para armazenar os dados de medição resultantes de um processo de negócio em um único *data mart*. Como os dados de medição são esmagadoramente a maior parte de qualquer *data mart*, deve-se evitar a duplicação em vários locais da empresa.

Machado (2004, página 79) define Fato da seguinte maneira “Um fato é uma coleção de itens de dados, composta de dados de medidas e de contexto”.

Cada fato representa um item, uma transação ou um evento de negócio e é utilizado para analisar o processo de negócio de uma empresa. É tudo aquilo que reflete a evolução dos negócios do dia-a-dia de uma organização.

Ainda para Machado (2004, página 100), “Fato é tudo aquilo que pode ser representado por um valor aditivo, ou melhor, sem academicismos, por meio de valores numéricos. Esse conjunto de valores numéricos é denominado de métricas ou medidas simplesmente.”

A característica básica de um fato é que ele é representado por valores numéricos e implementado em tabelas denominadas tabelas de fato (*fact tables*).

Outra característica importante para identificar um fato é que ele é evolutivo, muda suas medidas com o tempo, podendo ser sempre questionado sobre essa evolução ao longo de um espaço de tempo.

Kimball (2002, página 22), faz algumas considerações importantes acerca das tabelas de fato:

- Os principais fatos são numéricos e aditivos, como volume de vendas em quantidade monetária.;
- A capacidade de adição é fundamental porque as aplicações de *data warehouse* raramente recuperam uma única linha da tabela de fatos. Estas aplicações podem recuperar até milhões de linhas de fatos por vez;
- Existem fatos que são semi-aditivos e também fatos que são não-aditivos, quando isto acontece os semi-aditivos podem ser adicionados apenas a algumas dimensões. Já os não-aditivos não podem ser adicionados a nenhuma dimensão, restando então realizar contagens ou médias para podermos resumir as linhas;
- Normalmente, os fatos são descritos como contendo valores contínuos principalmente para que possam servir de guia para ajudar o projetista a classificar o que é um atributo de fato e um atributo de dimensão;
- Um fato medido pode ser textual, mas isso raramente acontece. Na maioria dos casos, uma medida textual é uma descrição de algo e é obtida de uma lista discreta de valores;
- Não se armazena informações textuais redundantes em tabelas de fato. A menos que o texto seja exclusivo para qualquer linha na tabela de fato, ele pertence à tabela dimensão;
- Normalmente as tabelas de fatos representam 90% ou mais do espaço total consumido por um banco de dados dimensional;
- A tabela de fatos tende a ser extensa em número de linhas, mas com pouco número de colunas;
- As tabelas de fatos têm as seguintes granularidades: transação, instantâneo periódico e instantâneo acumulado;
- As tabelas de fatos de granularidade de transação são as mais comuns;
- Tabelas de fatos possuem duas ou mais chaves estrangeiras;
- Trata-se de Integridade Referencial, a capacidade de acessar a tabela de fato, por meio de chave estrangeira, percorrendo as tabelas dimensão associadas a ela.

“Normalmente, a chave primária da tabela de fatos propriamente dita é formada por um subconjunto das chaves estrangeiras. Essa chave costuma ser denominada chave composta ou concatenada. Toda tabela de fatos em um modelo dimensional possui uma chave

composta e, da mesma forma, toda tabela possui uma chave composta em uma tabela de fatos. Uma outra maneira de dizer isso é que, em um modelo dimensional, toda tabela que expressa uma relação muitos-para-muitos deve ser uma tabela de fatos. Todas as outras tabelas são tabelas de dimensão” Kimball (2002, página 23)

DIMENSÃO

Para Kimball (2002, página 24) “As tabelas de dimensão estão sempre acompanhando uma tabela de fatos. Elas contêm descritores textuais da empresa. Em um modelo dimensional bem projetado, as tabelas de dimensão possuem muitas colunas ou atributos. Estes atributos descrevem as linhas na tabela de dimensão. Deve-se incluir o maior número possível de descrições semelhantes a texto significativas”.

Para Machado (2004, página 80) dimensão “...conceitualmente são os elementos que participam de um fato, assunto de negócios...”.

São as possíveis formas de visualizar os dados, ou seja, são os ‘por’ dos dados: ‘por mês’, ‘por país’, ‘por produto’, ‘por região’, etc.

As dimensões determinam o contexto de um assunto de negócios, por exemplo, um banco de dados que analise as vendas de produtos. As dimensões que participam desse fato vendas de produtos comumente são:

- Tempo;
- Localização;
- Clientes;
- Vendedores;
- Cenários (realizados, projetados).

Dimensões normalmente não possuem atributos numéricos, pois são somente descritivas e classificatórias dos elementos que participam de um fato.

Uma dimensão pode conter muitos membros e que um membro de uma dimensão é um nome diferente utilizado para determinar a posição de um item de dado. Por exemplo, todas as ocorrências de ano, trimestre e mês fazem a dimensão tempo, e todas as cidades, estados e regiões fazem a dimensão geográfica. Machado (2004, página 80)

Com relação as tabelas dimensão, Kimball(2002, página 24), faz algumas considerações:

- É normal tabelas de dimensão com 50 a 100 atributos;

- As tabelas de dimensão tendem a ser relativamente simples no que diz respeito ao número de linhas (normalmente, bem menos que 1 milhão de linhas), mas contêm um número bem grande de colunas;
- Cada dimensão é definida por sua chave primária, que funciona como base da integridade referencial com qualquer tabela de fatos a que esteja associada;
- Os atributos dimensionais, desempenham um papel fundamental, funcionam como uma fonte primária de restrições de consultas, agrupamentos e rótulos de relatórios isto faz com que o *data warehouse* possa ser usado e compreendido;
- Em uma solicitação de consulta ou de relatório, os atributos são identificados pelas palavras *by* (por). Por exemplo, quando um usuário afirma que deseja ver as vendas em valores monetários por semana e por marca, as palavras semana e marca devem estar disponíveis como atributos de dimensões;
- A potência do *data warehouse* é diretamente proporcional à qualidade e à complexidade dos atributos de dimensões;
- Quanto mais tempo se dedica à inclusão de atributos com terminologia de negócio verbosa, melhor é o *data warehouse*. Da mesma forma, quanto mais tempo se dedica ao preenchimento dos valores em uma coluna de atributos, melhor é o *data warehouse*;
- Os melhores atributos são os textuais e os discretos. Os atributos devem ser formados por palavras reais e não por abreviações criptográficas. Como exemplos de atributos típicos da dimensão de um produto, podemos citar uma descrição resumida (que contém de 10 a 15 caracteres), uma descrição longa (que contém de 30 a 50 caracteres) , o nome de uma marca, o nome de uma categoria, o tipo de embalagem, o tamanho e várias outras características do produto;
- É prudente diminuir o uso de códigos nas tabelas de dimensão substituindo-os por atributos textuais mais verbosos;
- Traduzir os códigos operacionais em atributos das tabelas dimensionais, extraíndo o seu significado no mundo operacional é uma prática fundamental para o sucesso do *data warehouse*;
- As tabelas de dimensão representam relações hierárquicas na empresa. Em uma tabela de dimensão de produto, por exemplo, produtos são divididos por marcas e depois por categorias. Para cada linha na dimensão do produto, deve-se incluir

a descrição da marca e da categoria associada a cada produto. Neste caso o preço da redundância é aceitável por facilitar a utilização e o desempenho da consulta.

- Quando se armazena apenas um código, por exemplo código de marca de produto na dimensão de uma tabela e cria-se uma tabela de consulta de marcas separadas. Essa tabela pode ser denominada de *snowflake*;
- Geralmente, as tabelas de dimensão não possuem muita normalização. Elas costumam ser bem pequenas (menos de 10% dos requisitos totais de armazenamento de dados) Como as tabelas de dimensão normalmente são geometricamente menores que as de fatos, melhorar a eficiência de armazenamento normalizando ou criando-se um *snowflaking* produz praticamente nenhum impacto no tamanho total do banco de dados. Deve-se priorizar a simplicidade e acessibilidade, em detrimento de um ganho de espaço inexpressivo no banco de dados.

MEDIDAS

Machado (2004, página 81), explica que, “Medidas são os atributos numéricos que representam um fato, a performance de um indicador de negócios relativo às dimensões que participam desse fato.

Os números atuais são denominados de variáveis. Por exemplo, medidas são o valor em reais das vendas, o número de unidades de produtos vendidas, a quantidade em estoque, o custo de venda, entre outros.

Uma medida, é determinada pela combinação das dimensões que participam de um fato, e estão localizadas como atributos de um fato”.

A principal utilização de um *data mart* é para consultar dados históricos que estão normalmente sumarizados por períodos de tempo e as mais variadas combinações de classificação de uma informação. Mas geralmente o que se deseja ver são valores numéricos e sua evolução ou não, em um espaço de tempo, com cálculos de transformação desses dados.

Esses valores numéricos são denominados de medidas ou métricas. Explica Machado (2004, página 135).

Ainda Machado (2004, página 136), explica que, as medidas se classificam em dois tipos sempre:

- Valores aditivos, que são aqueles referentes aos fato sobre os quais podem ser aplicadas as operações de soma, subtração e média. Os valores, como, por

exemplo, “número de crimes” e “números por tipo de transplante”, representam valores aditivos.

- Valores não aditivos: referentes aos fatos que não podem ser manipulados livremente, como valores percentuais ou relativos. Na realidade representam os indicadores de desempenho do fato.

A modelagem dimensional nada mais é que a reunião de tabelas de *fatos* e tabelas de *dimensão* relacionadas em um único modelo.

Os modelos dimensionais são bastante flexíveis para acomodar mudanças. A estrutura previsível de um modelo dimensional suporta mudanças inesperadas no comportamento do usuário. Toda dimensão é equivalente; e todas as dimensões são pontos de entrada simetricamente iguais para a tabela de fatos.

Com os modelos dimensionais, podemos adicionar dimensões totalmente novas ao esquema desde que um único valor dessa dimensão esteja definido para cada linha de fato existente. Explica Kimball (2002, página 28)

AGREGADOS

Os valores agregados representam paradoxalmente uma solução e alguns problemas. Solução, porque estabelece a definição e a criação de tabelas prontas, trabalhadas e sumariadas em várias dimensões, e que facilitam os acessos aos dados e agilizam os processos decisórios. Problemas, pois agride de certa forma os processos canônicos de não- redundância, estabelecidos nos preceitos de projetos de Bancos de Dados desde a sua criação. Além disso, obviamente, gastam mais espaço, pois exigirão uma coleção de tabelas Fato ou Dimensão, agora dedicadas ao armazenamento de dados num estado já pré-processado. Barbieri (2001, página 109)

Kimball (2002, página 30) , aponta “os mitos que cercam a modelagem dimensional”:

Mito 1 Os *data marts* e modelos dimensionais são usados apenas para dados de resumo. **O Importante é:** disponibilizar dados mais detalhados que respondam aos questionamentos dos usuários. Conciliar granularidade com resumos.

Mito 2 Os *data marts* e os modelos dimensionais são soluções departamentais e não corporativas. **O importante é:** organizar os *data marts* com base nos processos de negócio e não na visão departamental da empresa.

Mito 3 os *data marts* e modelos dimensionais não são escalonáveis. **O importante é:** os fornecedores de DBMSs relacionais atualmente já incorporaram vários recursos para otimizarem a escalabilidade e o desempenho dos modelos dimensionais.

Mito 4 os *data marts* e modelos dimensionais são apropriados apenas quando existe um padrão de utilização previsível. **O importante é:** projetar *data marts* extremamente flexíveis e adaptáveis às mudanças, criando tabelas de fatos no nível mais granular possível. Modelos dimensionais que só produzem dados de resumo tendem a ser problemáticos.

Mito 5 os *data marts* e modelos dimensionais não podem ser integrados e, portanto, levam a soluções isoladas. **O importante é:** modelos dimensionais obedecerem à arquitetura de barramento dos *data marts*. Os bancos de dados da área de apresentação que não obedecem a arquitetura de barramento levam a soluções independentes, e por consequência, isoladas.

2.2.5 OLAP: APLICAÇÕES ANALÍTICAS

É o conjunto de ferramentas que possibilita efetuar a exploração dos dados de um *data warehouse*, Machado (2004, página 86).

2.2.5.1 FERRAMENTAS OLAP

Machado (2004, página 86), conceitua que as ferramentas OLAP são aplicações as quais os usuários finais têm acesso para extrair os dados de suas bases e construir os relatórios capazes de responder às suas questões gerenciais.

A funcionalidade de uma ferramenta OLAP é caracterizada pela análise multidimensional dinâmica dos dados, apoiando o usuário final nas suas atividades, tais como *slice and dice e drill*. Operações OLAP.

Segundo Machado (2004, página 86), existem várias operações usadas em OLAP para analisar dados: *drill, slice, dice e pivot*. Operações *drill*, utilizam a navegação nos dados, modificando o nível de granularidade da consulta, podem ser *down, up, across ou thought*. O *down* vai da granularidade mais baixa para a granularidade mais alta. O *up* vai da granularidade mais alta para a granularidade mais baixa. O *across* quando o usuário pula um nível intermediário dentro de uma mesma dimensão. E *thought* que ocorre quando o usuário passa de uma informação contida em uma dimensão para a mesma informação contida em outra dimensão.

Operações *slice and dice* são operações para realizar navegação por meio dos dados na visualização de um cubo.

Slice é a operação que corta o cubo, mas mantém a mesma perspectiva da visualização dos dados.

Dice ou rotação significa mudar a perspectiva de visão, girando o cubo e invertendo uma ou mais dimensões. É como se tivessemos realizado a extração de um “subcubo” ou a interseção de vários *slices*.

Pivot é o ângulo pelo qual os dados são vistos ou trocados. Na prática, corresponde à modificação da posição das dimensões em um gráfico ou troca de linhas por colunas em uma tabela.

Na realidade, explica Machado (2004, página 93), é uma característica de planilhas eletrônicas, a sua capacidade de realizar *slice*.

2.3 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO

Neste capítulo, podemos observar que os sistemas de informação surgiram como resposta as necessidades operacionais e gerenciais das empresas. A evolução destes sistemas resultou numa arquitetura – *business intelligence*, que reúne várias tecnologias e que, quando bem aplicada, pode auxiliar muito os executivos das empresas a tomar decisões estratégicas. Também, foi visto que um dos pontos importante de *business intelligence* é o processo de ETL

(Extração, Transformação e Carga) dos dados que serão transformados em repositório de informações: o *Data Warehouse*. Este por sua vez, é um banco de informações estruturado (não-volátil, agrupado por assuntos, integrado e variável em relação ao tempo) cujo objetivo é dar respostas aos questionamentos estratégicos. Para que o *data warehouse* cumpra seu papel, devemos executar uma boa modelagem dimensional, onde a granularidade é que irá proporcionar consultas em cima dos fatos e dimensões sumariados do negócio da empresa, por meio das operações de OLAP, que possibilitam a obtenção das informações do *data warehouse*, “ao gosto do freguês”.

Como parte deste projeto, os próximos capítulos abordarão a metodologia da pesquisa, modelagem, processo de ETL e a criação de visualizações e inferências.

3 METODOLOGIA

Silva e Menezes (2001, página 19) definem que “pesquisar significa, de forma bem simples, procurar respostas para indagações propostas”.

A pesquisa é um conjunto de ações, propostas para encontrar a solução para um problema, que se tem por base procedimentos racionais e sistemáticos, Silva e Menezes (2001, página 20)

Uma pesquisa pode ser classificada metodologicamente sob quatro aspectos: sua natureza; a abordagem do problema; seus objetivos; e seus procedimentos técnicos.

3.1 TIPO DE PESQUISA

Do ponto de vista de sua natureza, o presente projeto classifica-se como sendo uma pesquisa aplicada, pois trata de gerar conhecimento para aplicação prática e dirigida à solução de problemas específicos. Silva e Menezes (2001, página 20)

Já a abordagem do problema, deste projeto, coloca-o como uma pesquisa qualitativa, pois não requer o uso de métodos e técnicas estatísticas. O ambiente natural é a fonte direta para coleta de dados e o pesquisador é o instrumento-chave. [...] O processo e seu significado são os focos principais de abordagem.

O objetivo deste projeto é proporcionar maior familiaridade com o problema com vistas a torná-lo explícito. Envolve levantamento bibliográfico; entrevistas com pessoas que têm experiências práticas com o problema pesquisado e análise de exemplos que estimulem a compreensão. Com isto posto, o presente projeto sob o ponto de vista de seus objetivos lhe confere caráter de uma pesquisa exploratória.

E finalmente, do ponto de vista de procedimentos técnicos, este projeto classifica-se como sendo uma pesquisa bibliográfica, pois está sendo elaborado a partir de material já publicado, constituído principalmente de livros, artigos de periódicos e atualmente com material disponibilizado na Internet. Mas também assume um aspecto de estudo de caso, pois envolve o estudo aprofundado em uns poucos objetos de maneira que se permita o seu amplo e detalhado conhecimento.

3.2 ETAPAS METODOLÓGICAS

De maneira geral, este projeto propõe a construção de uma arquitetura de *Business Intelligence* que será aplicado às tabelas que constam no Portal de Acesso à Informação e Transparência – Tabela de Lotação de Pessoal do Tribunal de Justiça de Santa Catarina. Estas tabelas analíticas, são: TLP 1 – Unidades Judiciárias de Primeiro e Segundo Graus, TLP 2 - Unidades de Apoio Direto à Atividade Judicante e TLP 3 - Unidades de Apoio Indireto à Atividade Judicante, reguladas e definidas pela Resolução 219/2016 do Conselho Nacional de Justiça.

Serão extraídos dados relevantes (dados disponibilizados pelo site que estão disponíveis à Sociedade Catarinense), identificados pelo acadêmico, das tabelas em questão, que serão persistidos em banco de dados relacional e formarão as dimensões e fatos do modelo dimensional, anteriormente projetado.

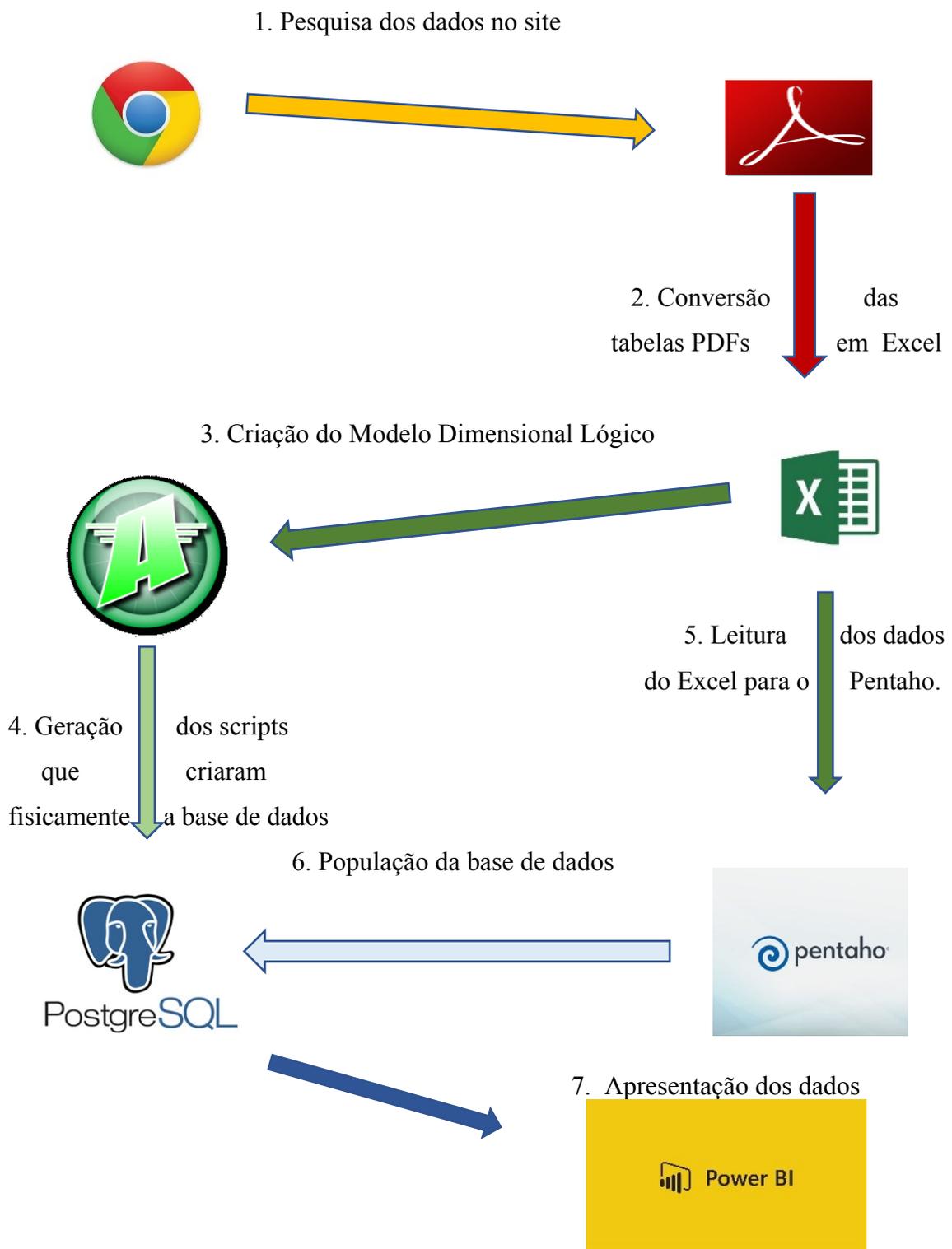
Na sequência, serão construídas as visualizações e inferências em cima do modelo dimensional projetado e persistido. Estas visualizações e inferências serão construídas a partir de uma ferramenta comercial, existente no mercado.

Estas tarefas representam as seguintes etapas:

1. Pesquisa dos dados no site;
2. Conversão das tabelas;
3. Criação do Modelo Dimensional Lógico;
4. Criação do Modelo Dimensional Físico;
5. População da base de dados;
6. Apresentação dos dados.

3.3 PROPOSTA DA SOLUÇÃO

Figura 4: Proposta da solução escolhida



Fonte: Autor (2018)

Abaixo uma descrição dos elementos da Proposta de Solução escolhida:

- Pesquisa dados no site – será utilizado o buscador Chrome para identificar as tabelas a serem manuseadas;
- Conversão das tabelas – será utilizado o Adobe Acrobat Reader para converter as tabelas de PDF para Excel – fará parte do processo de extração e transformação;
- Criação do Modelo Dimensional Lógico – será utilizado o SQL Power Architect;
- Criação do Modelo Dimensional Físico – serão utilizados scripts do SQL Power Architect para criação do modelo físico no Banco de Dados PostgreSQL;
- População da base de dados – será utilizado o Pentaho DataIntegration (Kettle) para popular a base de dados no PostgreSQL. Fará parte do processo de carga.
- Apresentação dos dados – será utilizado o Power BI para apresentação visual e inferências das informações.

3.4 DELIMITAÇÃO DO PROJETO

Este projeto está delimitado às tabelas disponíveis no site do Tribunal de Justiça de Santa Catarina e que constam no Portal de Acesso à Informação e Transparência – Tabela de Lotação de Pessoal, que são reguladas e definidas pela Resolução 219/2016 do Conselho Nacional de Justiça.

4 MODELAGEM

Neste capítulo será abordada a modelagem deste projeto. A modelagem divide-se em: requisitos (funcionais e não funcionais) e modelo dimensional.

4.1 REQUISITOS

Talvez o maior problema que enfrentamos no desenvolvimento de sistemas de software grandes e complexos seja o da engenharia de requisitos (Sommerville, 2007, página 76). No entanto, respeitada as devidas proporções, mesmo em pequenos sistemas, a definição de requisitos sempre é desafiadora. Ainda, segundo Sommerville (2007), os requisitos de um sistema, são descrições dos serviços fornecidos pelo sistema e as suas restrições operacionais. O termo requisito não é usado pela indústria de software de maneira consistente. Em alguns casos, um requisito é simplesmente uma declaração abstrata de alto nível..., nos afirma Sommerville, 2007.

Os requisitos dividem-se em 2 tipos: os funcionais e os não funcionais.

4.1.1 REQUISITOS FUNCIONAIS

Requisitos funcionais definem funcionalidades que devem ser criadas para permitir que usuários realizem suas tarefas ou para satisfazer requisitos de negócio, definindo o comportamento do sistema, expresso na forma de serviços ou funções.

Requisitos funcionais encontrados:

- RF 001: O sistema deve importar 3 tabelas do Portal do TJSC;
- RF 002: O sistema deve converter as tabelas em planilhas eletrônicas;

- RF 003: O sistema deve unificar as três planilhas eletrônicas;
- RF 004: O sistema deve acrescentar o semestre, o ano e o tipo de tabela;
- RF 005: O sistema deve dar carga no modelo dimensional na base de dados relacional;
- RF 006: O sistema deve permitir consultas visuais parametrizadas;

Tabela 1: descrição dos atributos dos requisitos funcionais

Identificador	Descrição
RF 001	O sistema deve importar 3 tabelas (em PDF), TLP 1, TLP 2 e TLP 3 do portal da transparência Tabela de Lotação de Pessoal (link Gestão de Pessoas - Acesso à Informação e transparência do TJSC - url: https://www.tjsc.jus.br/web/aceso-a-informacao-e-transparencia/gestao-de-pessoas)
RF 002	O sistema deve converter as tabelas TLP 1, TLP 2 e TLP 3 do portal da transparência Tabela de Lotação de Pessoal (link Gestão de Pessoas - Acesso à Informação e transparência do TJSC - url: https://www.tjsc.jus.br/web/aceso-a-informacao-e-transparencia/gestao-de-pessoas) que estão no formato PDF para XLS (Excel)
RF 003	O sistema deve juntar as três tabelas (em xls) em uma única tabela (xls)
RF 004	O sistema deve acrescentar o semestre, o ano e o tipo de tabela. Estas informações conforme o link acessado no portal da transparência Tabela de Lotação de Pessoal (link Gestão de Pessoas - Acesso à Informação e transparência do TJSC - url: https://www.tjsc.jus.br/web/aceso-a-informacao-e-transparencia/gestao-de-pessoas)
RF 005	O sistema deve dar carga no modelo dimensional na base de dados criada no Banco de Dados Relacional PostgreSQL
RF 006	O sistema deve permitir consultas visuais parametrizadas na ferramenta Power BI da Microsoft

Fonte: Autor (2018).

4.1.2 REQUISITOS NÃO-FUNCIONAIS

Os requisitos não funcionais, tratam de restrições, enfrentadas pelo sistema. Estas podem ser de timing, processo de desenvolvimento e padrões. Foram encontrados os seguintes requisitos não-funcionais:

Requisitos não-funcionais encontrados:

- RNF 001: O sistema deve utilizar para a carga dos dados o Pentaho Integration;
- RNF 002: O sistema deve utilizar o PostgreSQL como Banco de Dados;
- RNF 003: O sistema deve utilizar como ferramenta visual o Power B.I.

Tabela 2: descrição dos atributos dos requisitos não-funcionais

Identificador	Descrição
RNF 001	O sistema deve utilizar para a carga dos dados o Pentaho Integration.
RNF 002	O sistema deve utilizar o PostgreSQL como Banco de Dados.
RNF 003	O sistema deve utilizar como ferramenta visual o Power B.I.

Fonte: Autor (2018)

4.2 MODELAGEM DIMENSIONAL

A modelagem dimensional é a técnica de projeto que conduz os dados a formação de um modelo onde a informação reside na interseção de várias dimensões ... (Barbieri, 2001).

O modelo dimensional, será composto das seguintes dimensões:

Dimensão Tempo, Dimensão Geografia, Dimensão TLP. Estas dimensões estarão interagindo com o Fato Lotação.

4.2.1 DIMENSÃO TLP

Dimensão TLP (DI_TLP), é uma tabela que armazena, o grau de jurisdição da unidade, o nome da unidade, o tipo da unidade e a descrição a unidade.

Figura 5: Dimensão TLP

DI_TLP
SEQ_TLP: INTEGER NOT NULL [PK]
GRAU_JURISDICA0: VARCHAR(10) NOT NULL
NME_UNIDADE_JUDICIARIA: VARCHAR(80) NO
TPO_UNIDADE_JUDICIARIA: VARCHAR(60) NO
DSC_UNIDADE_JUDICIARIA: VARCHAR(200) NQ

Fonte: Autor (2018)

Tabela 3: descrição dos atributos da dimensão TLP

Atributo	Descrição do atributo
GRAU_JURISDICA0	Os graus podem ser: 1 (primeiro grau), 2 (segundo grau), ambos (quando a unidade serve a ambos os graus);
NME_UNIDADE_JUDICIARIA	Nome da Unidade Judiciária, está ligada ao tipo de tabela no Portal da Transparência do TJSC, varia entre: TLP 1 – Tabela de Lotação de Pessoal das Unidades Judiciárias de Primeiro e Segundo Graus TLP 2 – Tabela de Lotação de Pessoal das Unidades de Apoio Direto à Atividade Judicante TLP 3 – Tabela de Lotação de Pessoal das Unidades de Apoio Indireto à Atividade Judicante Este dado consta no link de consultas das tabelas do TJSC e foi levado para dentro do modelo dimensional como um campo.
TPO_UNIDADE_JUDICIARIA	Tipo de Unidade, pode ser Vara, Juizado Especial, Turma Recursal, Gab. Des. (Gabinete de Desembargador) e Outros Tipo de Unidade, pode ser

	Vara, Juizado Especial, Turma Recursal, Gab. Des. (Gabinete de Desembargador) e Outros.
DSC_UNIDADE_JUDICIARIA	Descrição da Unidade, abrange um amplo espectro.

Fonte: Autor (2018)

4.2.2 DIMENSÃO GEOGRAFIA

Dimensão Geografia (DI_GEOGRAFIA), é uma tabela que armazena, o nome do município onde está localizada a Unidade (em referência a uma linha da tabela) em questão. E também a UF (unidade da federação ou estado) em que está localizada a Unidade (em referência a uma linha da tabela). Por se tratar do Portal da Transparência do Tribunal de Justiça de Santa Catarina, o valor é fixo em SC.

Figura 6: Dimensão Geografia

DI_GEOGRAFIA	
SEQ_GEOGRAFIA:	INTEGER NOT NULL
NME_MUNICIPIO:	VARCHAR(6) NOT NULL
SGL_ESTADO:	VARCHAR(2) NOT NULL

Fonte: Autor (2018)

Tabela 4: descrição dos atributos da dimensão Geografia

Atributo	Descrição do atributo
NME_MUNICIPIO	nome do município onde esta localizada a Unidade.
SGL_ESTADO	UF, em se tratando do TJSC, sempre será SC.

Fonte: Autor (2018)

4.2.3 DIMENSÃO TEMPO

Dimensão Tempo (DI_TEMPO), é uma tabela que armazena, o ano e o semestre. O ano no forma NNNN (exemplo: 2017) e o semestre no formato N (1 para primeiro e 2 para o segundo semestre). Tanto ano, como semestre são inferidos pelo link da Tabela de Gestão de Pessoas, no Portal da Transparência do TJSC.

Figura 7: Dimensão Tempo

DI_TEMPO
SEQ_TEMPO: INTEGER NOT NULL
ANO: INTEGER NOT NULL
SEMESTRE: VARCHAR(20) NOT N

Fonte: Autor (2018)

Tabela 5 – Descrição dos atributos da tabela dimensão Tempo

Atributo	Descrição do atributo
ANO	pode ser 2017 ou 2018. O ano consta no link da consulta das tabelas no site do TJSC e foi levado para dentro do modelo dimensional como um campo.
SEMESTRE	O semestre é inferido pelo mês que consta no link da consulta das tabelas no site do TJSC e foi levado para dentro do modelo dimensional como um campo.

Fonte: Autor (2018)

4.2.4 TABELA FATO LOCAÇÃO

A tabela fato de lotação (FT_LOCACAO), armazena, os seguintes campos:

LP, lotação paradigma (número, que segundo a Resolução 219/2016 do CNJ), define o número ideal de servidores naquela unidade, LR Lotação Real dos Efetivos, LR_I, lotação real dos que ingressaram por cessão ou requisição, LR_SV, lotação Real dos servidores sem vínculo, CC, cargos em comissão, FC, função em confiança.

Figura 8: Fato Locação

FT_LOCACAO
SEQ_TLP: INTEGER NOT NULL [PFK]
SEQ_GEOGRAFIA: INTEGER NOT NULL
SEQ_TEMPO: INTEGER NOT NULL [PF
LP: INTEGER NOT NULL
LR_EFET: INTEGER NOT NULL
LR_I: INTEGER NOT NULL
LR_SV: INTEGER NOT NULL
CC: INTEGER NOT NULL
FC: INTEGER NOT NULL

Fonte: Autor (2018)

Tabela 6 – Descrição dos atributos da tabela fato Locação

Atributo	Descrição do atributo
LP	Lotação Paradigma: quantitativo mínimo de servidores das unidades judiciárias de primeiro e de segundo grau, de acordo com a Resolução 219 do CNJ Conselho Nacional de Justiça.
LR_EFET	Lotação Real dos Efetivos.
LR_I	Lotação Real dos que ingressaram por cessão ou requisição.
LR_SV	Lotação Real dos Servidores sem Vínculo
CC	Cargos em Comissão
FC	Funções de Confiança

Fonte: Autor (2018)

4.3 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO

Nesse capítulo foram apresentados os requisitos (funcionais e não funcionais) levantados a partir do acesso ao Portal da Transparência Gestão de Pessoas Tabelas de Lotação do Tribunal de Justiça de Santa Catarina. Também foi apresentado o Modelo Dimensional, formando com os requisitos a modelagem do nosso B.I. (*Business Intelligence*). O próximo capítulo trará do desenvolvimento do B.I.

5 DESENVOLVIMENTO

Neste capítulo, serão abordadas as etapas realizadas, as tecnologias utilizadas e ferramentas envolvidas na criação da arquitetura de *Business Intelligence*.

5.1 TECNOLOGIAS E FERRAMENTAS

Foram diversas ferramentas utilizadas neste projeto. Ei-las:

5.1.1 GOOGLE CHROME

O Chrome é um navegador de internet muito popular atualmente, tendo como concorrentes, entre outros o Mozilla FireFox e Internet Explorer, da Microsoft. Junto com um buscador (também de mesmo nome e também da Google) aumenta em muito a produtividade. Neste projeto, foi utilizado para encontrar as tabelas do Acesso à informação e Transparência – Gestão de Pessoas – Tabela de Lotação de Pessoal (TLP 1 – Unidades judiciárias de Primeiro e Segundo Grau; TLP 2 – Demais Unidade de Apoio Direto à Atividade Judicante e TLP 3 – Unidades de Apoio Indireto à Atividade Judicante) do Portal da Transparência do Tribunal de Justiça de Santa Catarina.

5.1.2 ADOBE ACROBAT READER

Importante gerador e leitor de arquivos no formato PDF, sigla inglesa para *Portable Document Format* (Formato Portátil de Documento) criado pela Adobe Systems. Neste projeto,

o Adobe Acrobat Reader, foi utilizado para ler, baixar (download) e converter (opção Exportar PDF do aplicativo mobile do Adobe Acrobat Reader) as tabelas do Acesso à informação e Transparência – Gestão de Pessoas – Tabela de Lotação de Pessoal (TLP 1 – Unidades judiciárias de Primeiro e Segundo Grau; TLP 2 – Demais Unidade de Apoio Direto à Atividade Judicante e TLP 3 – Unidades de Apoio Indireto à Atividade Judicante) do Portal da Transparência do Tribunal de Justiça de Santa Catarina. A utilização do Adobe Acrobat Reader para a conversão dos arquivos PDF em arquivos XLS (arquivos da planilha eletrônica do Microsoft Excel) teve uma taxa altíssima de conversão, porém não alcançou 100%. Diversos aplicativos de conversão PDF para XLS no entanto, foram testados, o melhor resultado foi obtido com o Adobe na versão paga.

5.1.3 MICROSOFT EXCEL

Planilha eletrônica desenvolvida pela Microsoft Corporation e pertencente ao grupo de aplicativos do pacote Office da Microsoft Corporation. O Excel é uma planilha eletrônica amplamente utilizada nos dias de hoje, e transformou-se numa poderosa ferramenta de produtividade, traz nas suas funcionalidades, entre outras, a possibilidade de criação de gráficos de fácil visualização. Neste projeto, foi utilizada para editar linhas e colunas das tabelas convertidas pelo Adobe Acrobat Reader (não houve 100% de conversão), organizar linhas e colunas e também criar novas colunas como SEMESTRE, ANO e o nome do tipo de tabela. Também, foi utilizada para unificar todas as tabelas do Acesso à informação e Transparência – Gestão de Pessoas – Tabela de Lotação de Pessoal (TLP 1 – Unidades judiciárias de Primeiro e Segundo Grau; TLP 2 – Demais Unidade de Apoio Direto à Atividade Judicante e TLP 3 – Unidades de Apoio Indireto à Atividade Judicante) do Portal da Transparência do Tribunal de Justiça de Santa Catarina.

O Excel foi uma importante ferramenta na etapa de Transformação do processo de ETL (vide capítulo 2).

5.1.4 SQL POWER ARCHITECT

Ferramenta para a criação de modelo de dados, que permite trabalhar de forma gráfica na modelagem dimensional do *Business Intelligence*. Neste projeto, o SQL Power Architect auxiliou na criação das tabelas de dimensão DI_TEMPO, DI_TLP e DI_TLP e na tabela fato FT_LOCACAO do nosso modelo dimensional. Algumas de suas funcionalidades são: Boa interface visual com diagrama E-R e OLAP; Acesso a banco de dados via JDBC; Conexão a múltiplas fontes de dados simultaneamente; Comparação de modelos de dados com identificação de diferenças e geração de scripts de sincronização; Suporte a “Drag-and-drop”, arrastar e soltar tabelas;

5.1.5 POSTGRESQL

É um Sistema Gerenciador de Banco de dados (objeto-relacional), de código aberto, ou seja, licenciamento livre o que confere a ele uma ampla fatia do mercado. Neste projeto, o PostgreSQL foi utilizado na criação das entidades para persistência dos dados do modelo dimensional definido no SQL Power Architect. Algumas de suas características são: Consultas complexas, Integridade transacional, Controle de concorrência multi-versão, Suporte ao modelo híbrido objeto-relacional entre outras.

5.1.6 PENTAHO DATA INTEGRATION (KETTLE)

O Pentaho Data Integration (Kettle) é uma poderosa ferramentas para transformações de dados (ETLs). Entre suas funcionalidades estão : migração de dados, conversões de sistemas ou carga de bancos de dados, transformação de dados, além de tarefas avançadas como

transferência de arquivos, envios de e-mails, integração com webservices e programações com scripts com schedules.

Neste projeto, ele foi utilizado para carga do banco de dados.

5.1.7 POWER BI

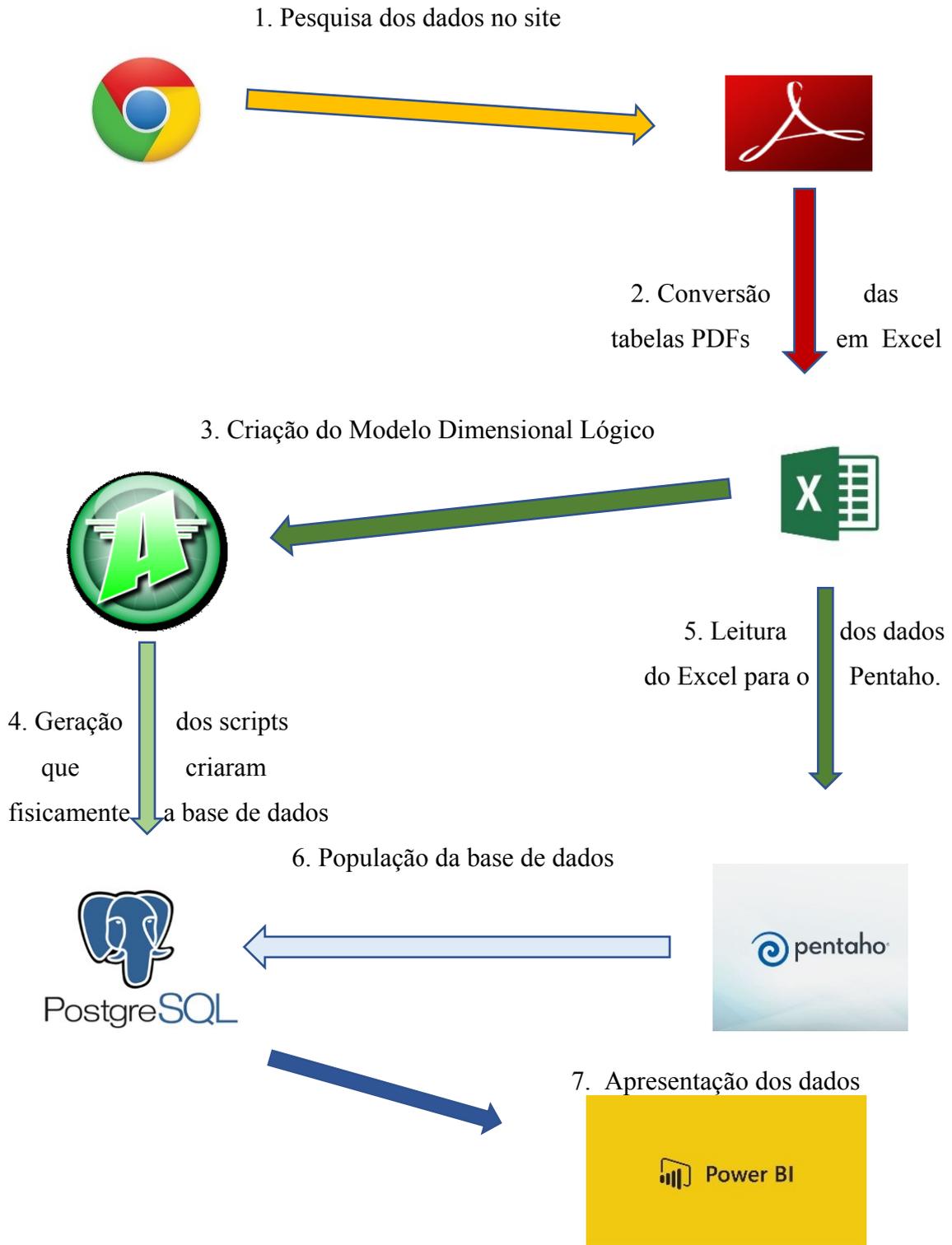
Uma poderosa ferramenta da Microsoft que une criação fácil com ampla visualização, dos dados tabelados. Reúne suplementos de BI encontrados no Autoatendimento do Excel Pro Plus (consulta poderosa com um poderoso pivô e uma, também, poderosa visualização). Neste projeto, foi utilizado na tarefa de interface com o usuário, fornecendo a este, uma excelente possibilidade de análise e inferências diversas.

5.2 ARQUITETURA DO BI.

Nesta e nas próximas sessões, será apresentada a arquitetura do BI. Os processos de Extração, Transformação e Carga; A Modelagem Dimensional; A Carga no Banco de Dados e a visualização e inferências do BI.

5.2.1 O PROCESSO

Figura 9: O Processo de Desenvolvimento (idem a figura 4 do capítulo 3)

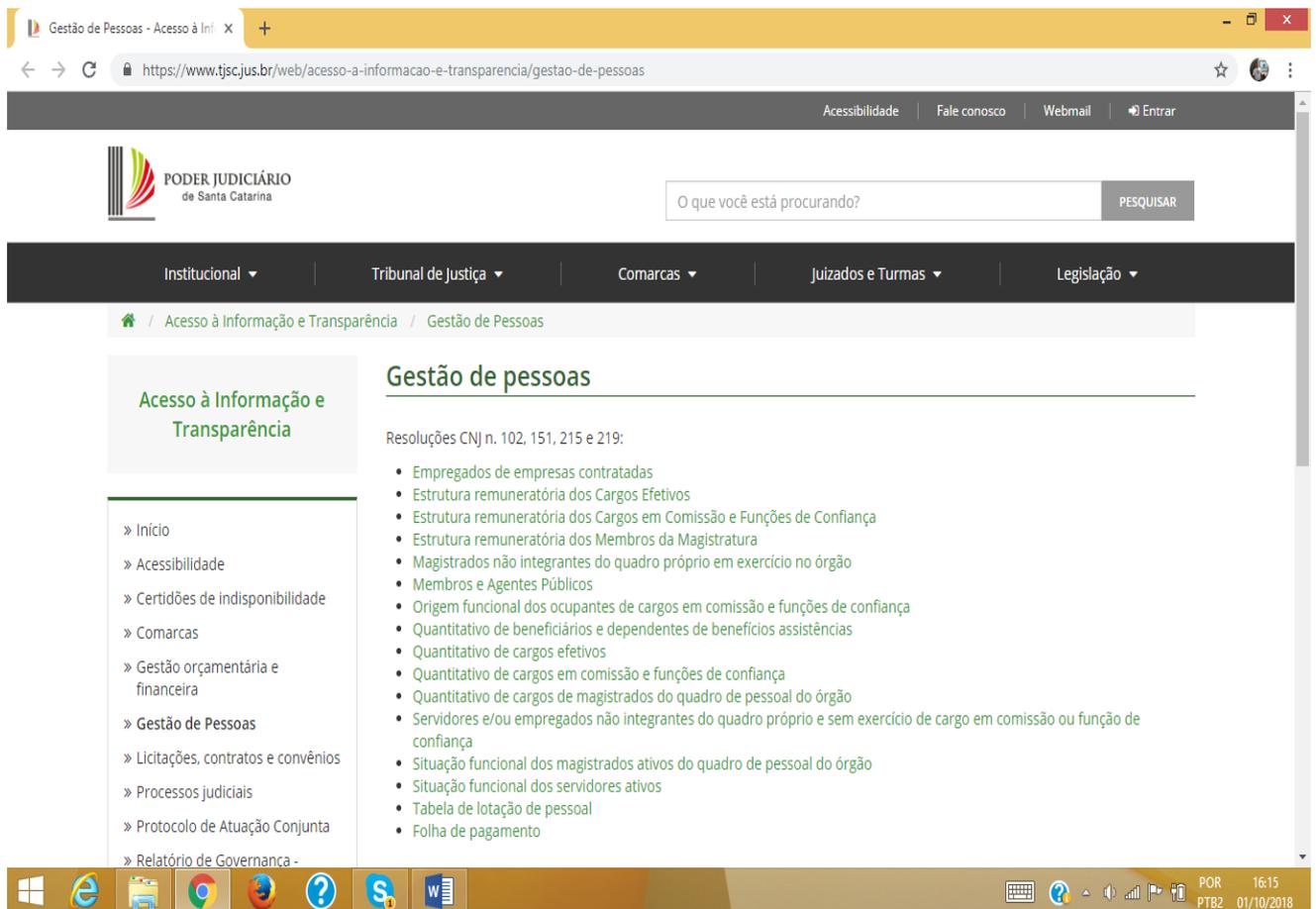


Fonte: Autor (2018)

5.2.2 Download das tabelas em PDF.

Acessar o Portal de Acesso à Informação e Transparência do Tribunal de Justiça de Santa Catarina, Gestão de Pessoas.

Figura 10: Site Portal de Acesso à Informação e Transparência do TJSC



Fonte: www.tjsc.jus.br/web/aceso-a-informacao-e-transparencia/gestao-de-pessoas

Acessar a Tabela de Lotação de Pessoal, que se abre em 3 opções: 2018 – janeiro, 2017 – julho e 2017 – janeiro (opções disponíveis em setembro de 2018)

Figura 11: Tabela de lotação de pessoal no Portal de Acesso à Informação e Transparência do TJSC – opção dos semestres

The screenshot shows a web browser window with the URL <https://www.tjsc.jus.br/web/aceso-a-informacao-e-transparencia/tabela-de-lotacao-de-pessoal>. The page header includes the logo of the Poder Judiciário de Santa Catarina and a search bar. The main navigation menu contains: Institucional, Tribunal de Justiça, Comarcas, Julzados e Turmas, and Legislação. The breadcrumb trail is: Acesso à Informação e Transparência / Gestão de Pessoas / Tabela de lotação de pessoal. The main content area is titled 'Tabela de lotação de pessoal' and features a list of semesters with expandable options:

Semestre	Ação
2018 - Julho	>
2018 - Janeiro	>
2017 - Julho	>
2017 - Janeiro	>

Buttons for 'ABRIR TODOS OS ITENS' are located at the top right and bottom right of the list. A sidebar on the left provides navigation under 'Acesso à Informação e Transparência' with links to: Início, Acessibilidade, Certidões de indisponibilidade, Comarcas, Gestão orçamentária e financeira, Gestão de Pessoas, Licitações, contratos e convênios, Processos judiciais, Protocolo de Atuação Conjunta, and Relatório de Governança.

Fonte: www.tjsc.jus.br/web/aceso-a-informacao-e-transparencia/tabela-de-lotacao-de-pessoal

“Clicar” na seta à direita de cada ano e mês. Aparecerá os 3 tipos de tabelas disponíveis para consulta e download. Os tipos são: (TLP 1 – Unidades judiciárias de Primeiro e Segundo Grau; TLP 2 – Demais Unidade de Apoio Direto à Atividade Judicante e TLP 3 – Unidades de Apoio Indireto à Atividade Judicante).

Figura 12 : Os 3 tipos de tabelas (TLP 1, TLP 2 e TLP 3) disponíveis para consulta e download no Portal de Acesso à Informação e Transparência no site do TJSC – opção tabelas

The screenshot shows the website interface for 'Tabela de lotação de pessoal'. The main content area is titled 'Tabela de lotação de pessoal' and features three expandable sections for different periods:

- 2018 - Janeiro:**
 - TLP 1 - Unidades Judiciárias de Primeiro e Segundo Grau
 - TLP 2 - Demais Unidades de Apoio Direto à Atividade Judiciante
 - TLP 3 - Unidades de Apoio Indireto à Atividade Judiciante
- 2017 - Julho:**
 - TLP 1 - Unidades Judiciárias de Primeiro e Segundo Grau
 - TLP 2 - Demais Unidades de Apoio Direto à Atividade Judiciante
 - TLP 3 - Unidades de Apoio Indireto à Atividade Judiciante
- 2017 - Janeiro:**
 - TLP 1 - Unidades Judiciárias de Primeiro e Segundo Grau

A sidebar on the left provides navigation options under 'Acesso à Informação e Transparência', including links to 'Início', 'Acessibilidade', 'Certidões de indisponibilidade', 'Comarcas', 'Gestão orçamentária e financeira', 'Gestão de Pessoas', 'Licitações, contratos e convênios', 'Processos judiciais', and 'Protocolo de Atuação Conjunta'.

Fonte: www.tjsc.jus.br/web/aceso-a-informacao-e-transparencia/tabela-de-lotacao-de-pessoal

Escolher qual o tipo de tabela a consultar.

Figura 13: A tabela escolhida para exemplo foi a TLP 1 - Unidades judiciárias de Primeiro e Segundo Grau

The document viewer displays the following content:

TLP 1 – Tabela de Lotação de Pessoal das Unidades Judiciárias de Primeiro e Segundo Grau

Grau – Grau de Jurisdição
 Tipo – Tipo de Unidade Judiciária
 Dsc_Unidade – Descrição da Unidade Judiciária
 UF – Unidade Federativa onde está localizada cada unidade judiciária
 Munic – Município
 LP – Lotação Paradigma
 LR_Efet – Lotação Real dos Efetivos
 LR_J – Lotação Real dos que ingressaram por cessão ou requisição
 LR_SV – Lotação Real dos Servidores sem Vínculo

Grau	Tipo	Dsc_Unidade	UF	Munic	LP	Lotação Real (LR)	LR_Efet	LR_J	LR_SV
2ª	Gab. Desembargador	Gabinete da Desembargadora Bettina Maria Maresch de Moura	SC	FLORIANOPOLIS	7	6	1		5
2ª	Gab. Desembargador	Gabinete da Desembargadora Cinthia Beatriz da Silva Bittencourt Schaefer	SC	FLORIANOPOLIS	7	6	1		5
2ª	Gab. Desembargador	Gabinete da Desembargadora Cibúdia Lambert de Faria	SC	FLORIANOPOLIS	7	7	2		5
2ª	Gab. Desembargador	Gabinete da Desembargadora Denise de Souza Luiz Francoski	SC	FLORIANOPOLIS	7	6	1		5
2ª	Gab. Desembargador	Gabinete da Desembargadora Denise Valpato	SC	FLORIANOPOLIS	7	6	1		5
2ª	Gab. Desembargador	Gabinete da Desembargadora Hídemar Meneguzzi de Carvalho	SC	FLORIANOPOLIS	7	6	2		4
2ª	Gab. Desembargador	Gabinete da Desembargadora Janice Goulart Garcia Ubiatti	SC	FLORIANOPOLIS	7	6	2		4
2ª	Gab. Desembargador	Gabinete da Desembargadora Maria do Rocio Luz Santa Rita	SC	FLORIANOPOLIS	7	9	3		6
2ª	Gab. Desembargador	Gabinete da Desembargadora Rejane Andersen	SC	FLORIANOPOLIS	7	9	3		4
2ª	Gab. Desembargador	Gabinete da Desembargadora Rosane Portella Wolff	SC	FLORIANOPOLIS	7	5	1		4
2ª	Gab. Desembargador	Gabinete da Desembargadora Salete Silva Sommariva	SC	FLORIANOPOLIS	7	8	4		4
2ª	Gab. Desembargador	Gabinete da Desembargadora Sônia Maria Schmitz	SC	FLORIANOPOLIS	7	8	2		6
2ª	Gab. Desembargador	Gabinete da Desembargadora Soraya Nunes Lima	SC	FLORIANOPOLIS	7	7	2		5
2ª	Gab. Desembargador	Gabinete da Desembargadora Vera Lúcia Ferreira Copetti	SC	FLORIANOPOLIS	7	6	2		4

Fonte: www.tjsc.jus.br/documents/72312/802964/TLP+1+-+Tabela+de+Lota%C3%A7%C3%A3o+de+Pessoal+das+Unidades+Judici%C3%A1rias+de+Primeiro+e+Segundo+Graus+-+Abril+2018/9d584de4-50cc-471a-8eed-8f1a0285f892

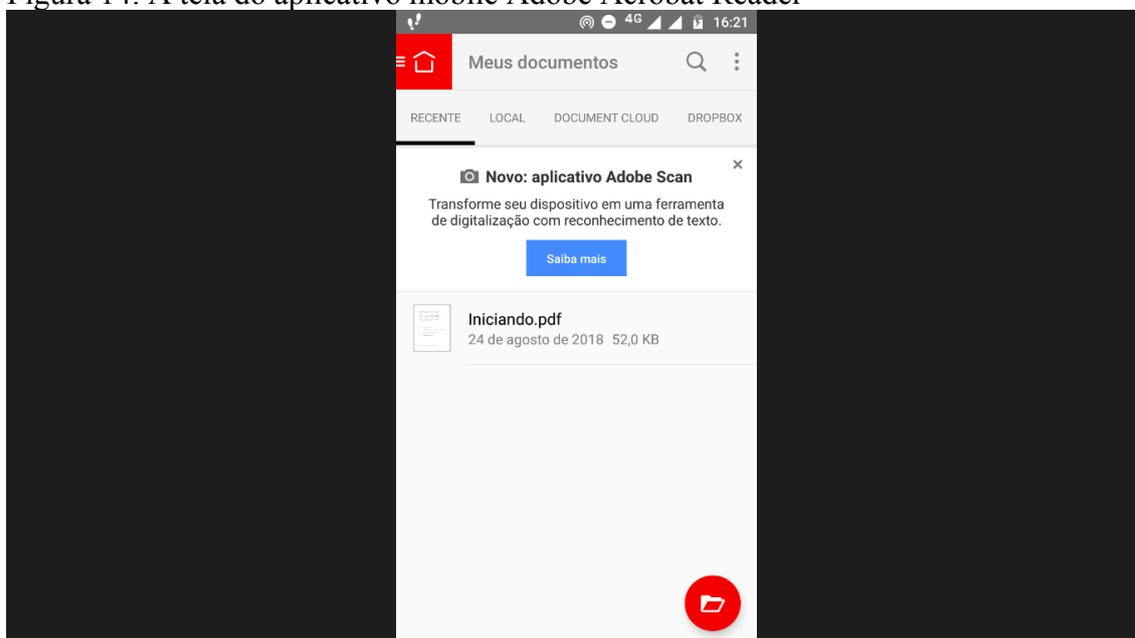
Com a tabela aberta, realizar o download. Por uma questão de produtividade, optar por baixar o PDF em um dispositivo mobile, o Adobe Acrobat Reader é uma boa opção.

5.2.3 Conversão das tabelas em PDF para Excel.

No dispositivo mobile, com o Adobe Acrobat Reader aberto executar os seguintes passos:

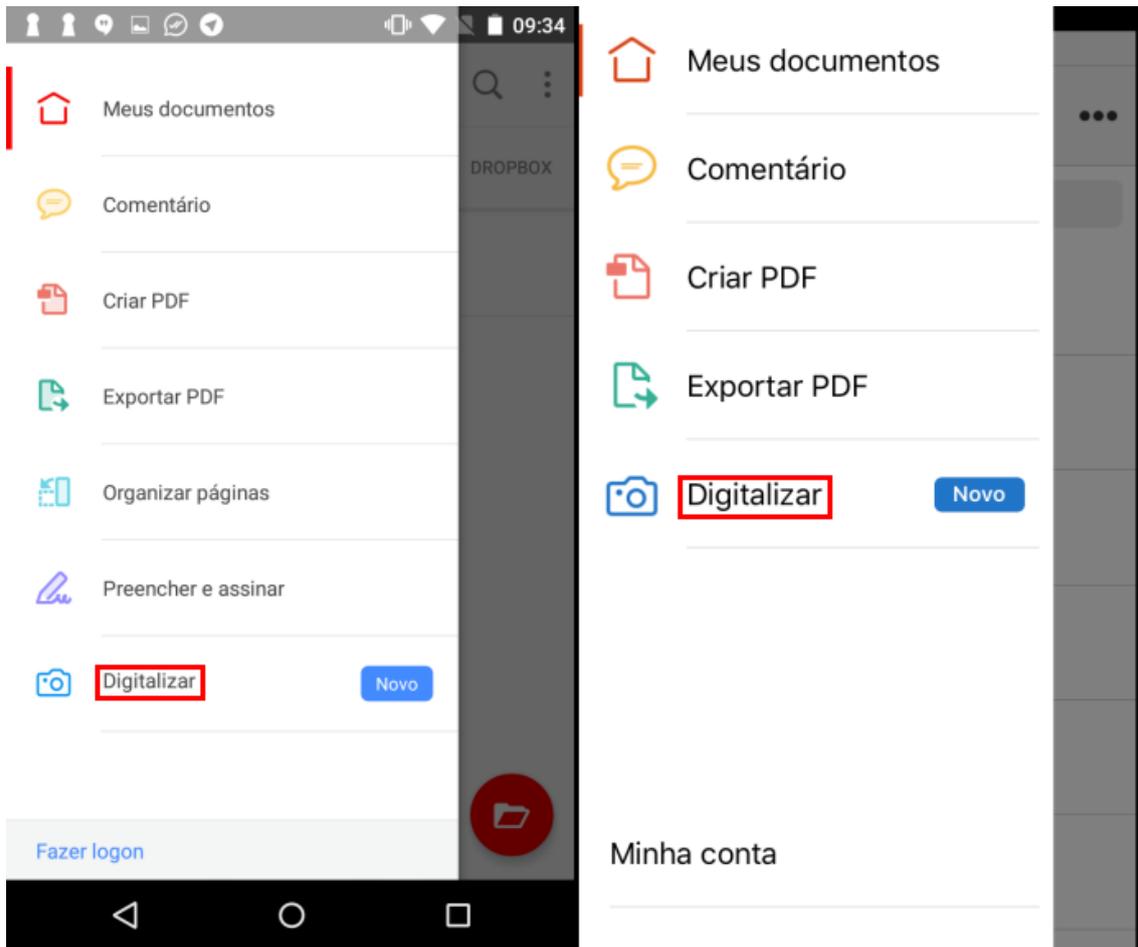
Acessar o Adobe Acrobat Reader.

Figura 14: A tela do aplicativo mobile Adobe Acrobat Reader



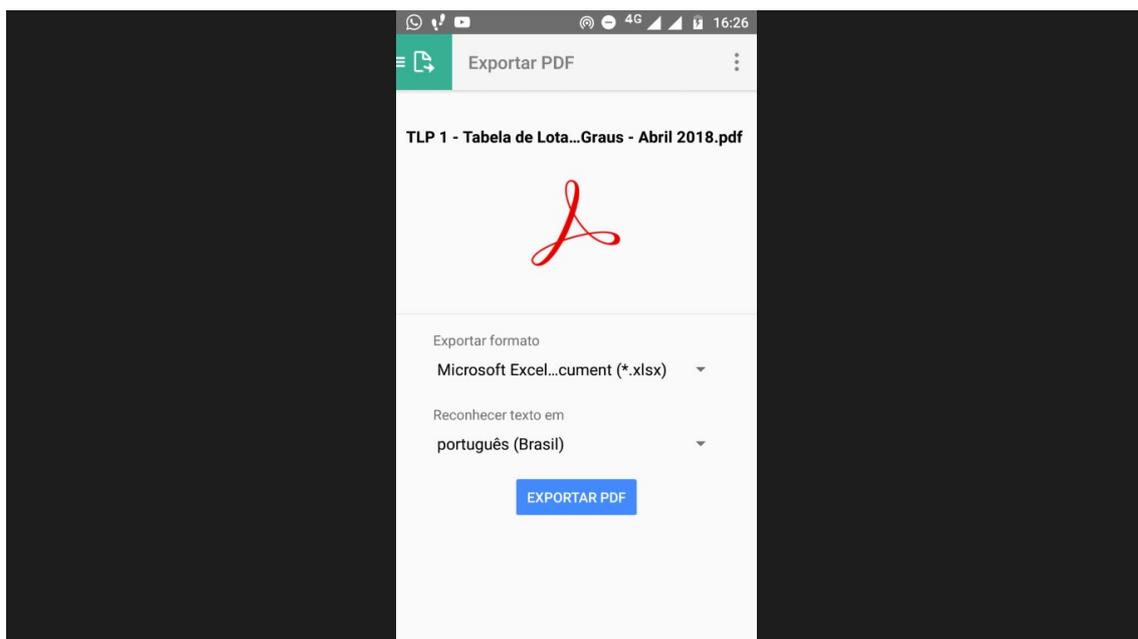
Fonte: Autor (2018)

Figura 15: Foi escolhida a opção Exportar PDF do aplicativo mobile Adobe Acrobat Reader



Fonte: Autor (2018)

Figura 16: Foi escolhido o tipo de arquivo de destino, no nosso caso foi XLSX



Fonte: Autor (2018)

5.2.4 Edição das planilhas em Excel.

Depois da exportação finalizada começar a edição das planilhas, a serem editadas para correção de dados, incluir e deslocar algumas colunas e juntar todas as 3 planilhas (TLP 1, TLP 2 e TLP 3) em uma única planilha (TLP).

Figura 17: Edição para correção e unificação das planilhas. Planilhas corrigidas e unificadas

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Gra	Tipo	Dsc_Unidade	UF	Munic	LP	LR_Efet	LR_I	LR_SV	Outros	CC	FC	SEMESTRE	ANO	NME
2	1	outros	CGJ/JC - NUCLEO IV	sc	Florianópolis	0,00	0	0	4	1	24	0	1	2017	TLP 3
3	2	outros	PRES.TJ - GABINETE DA PRESIDENCIA	sc	Florianópolis	0	9	0	9	1	24	0	1	2018	TLP 2
4	2	outros	PRES.TJ - GABINETE DA PRESIDENCIA	sc	Florianópolis	0,00	6	3	13	13	18	0	1	2017	TLP 2
5	0	outros	CGJ/JC - NUCLEO IV	sc	Florianópolis	0	15	0	1	0	16	0	1	2018	TLP 3
6	2	outros	DCDP - ASSESSORIA DE CADASTRAMENTO	sc	Florianópolis	0,00	17	0	0	0	16	0	1	2017	TLP 2
7	2	outros	DCDP - ASSESSORIA DE CADASTRAMENTO	sc	Florianópolis	0	21	0	0	0	16	0	1	2018	TLP 2
8	0	outros	CGJ - ASSESSORIA CORREICIONAL	sc	Florianópolis	0	3	0	0	0	14	0	1	2018	TLP 3
9	2	outros	GD - DES. JOSE ANTONIO TORRES MARQUES	sc	Florianópolis	0,00	2	0	12	12	13	0	1	2017	TLP 2
10	2	outros	GD - DES. RICARDO OROFINO DA LUZ FONTES	sc	Florianópolis	0,00	2	0	11	11	13	0	1	2017	TLP 2
11	2	outros	GD - DES. JOSE ANTONIO TORRES MARQUES	sc	Florianópolis	0	2	0	3	2	12	0	1	2018	TLP 2
12	2	outros	GD - DES. RICARDO OROFINO DA LUZ FONTES	sc	Florianópolis	0	0	0	5	2	12	0	1	2018	TLP 2
13	2	outros	1A. VICE-PRESIDENCIA	sc	Florianópolis	0,00	5	0	6	6	10	0	1	2017	TLP 2
14	2	outros	1A. VICE-PRESIDENCIA	sc	Florianópolis	0	3	0	3	2	10	0	1	2018	TLP 2
15	2	outros	GD - DES. ALEXANDRE DIVANENKO	sc	Florianópolis	0,00	3	0	7	7	9	0	1	2017	TLP 2
16	2	outros	GD - DES. ALEXANDRE DIVANENKO	sc	Florianópolis	0	3	0	6	2	9	0	1	2018	TLP 2
17	2	outros	GD - DES. RODOLFO CEZAR RIBEIRO DA SILVA TRIDAPALLI	sc	Florianópolis	0	2	0	5	2	9	0	1	2018	TLP 2
18	Amb	outros	GP - AUDITORIA INTERNA	sc	Florianópolis	0,00	0	1	0	0	9	0	1	2017	TLP 3
19	2	outros	02A. VICE-PRESIDENCIA	sc	Florianópolis	0,00	2	0	6	6	8	0	1	2017	TLP 2
20	2	outros	02A. VICE-PRESIDENCIA	sc	Florianópolis	0	3	0	3	0	8	0	1	2018	TLP 2

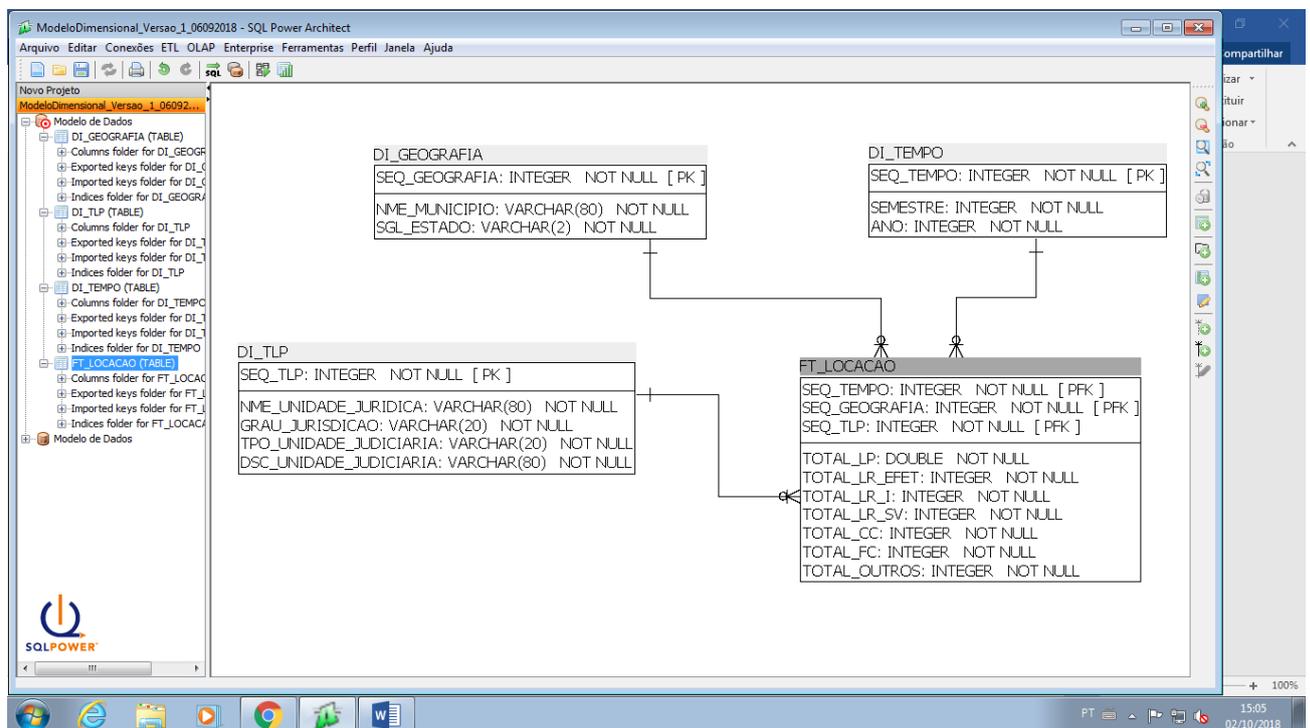
Fonte: Autor (2018)

5.2.5 O modelo dimensional

Para criação foi utilizada o SQL Power Architect, uma ferramenta poderosa que auxilia na criação do modelo dimensional lógico. Primeiro, foi analisada a planilha do Excel unificada, a partir daí, saltaram aos olhos a dimensão Tempo, composta por semestre e ano. Semestre, representa os seis meses anteriores ao próprio mês de referência (janeiro ou julho), por exemplo: janeiro de 2018, representa a contabilização dos cargos desde de 1º. agosto de 2017 a 31 de dezembro de 2017, conforme o Art. 15 da Resolução 219/2016. Ano é uma referência ao ano

comercial. Pode ser corrente ou anterior e junto com o semestre, conferem um aspecto único à contabilização do período. Outra Dimensão que chamou a atenção imediatamente foi a dimensão Geografia. Nesta dimensão as lotações são agrupadas pelos municípios aos quais pertencem. Nesta dimensão, a granularidade foi o fator preponderante. Logo, em seguida, foi definida a dimensão TLP, aqui também, a granularidade foi definitiva para a dimensão. Nossa tabela fato, contém todas as medidas para as inferências que são propostas, e completa nosso modelo dimensional lógico da arquitetura de BI.

Figura 18: Modelo Dimensional Lógico



Fonte: Autor (2018)

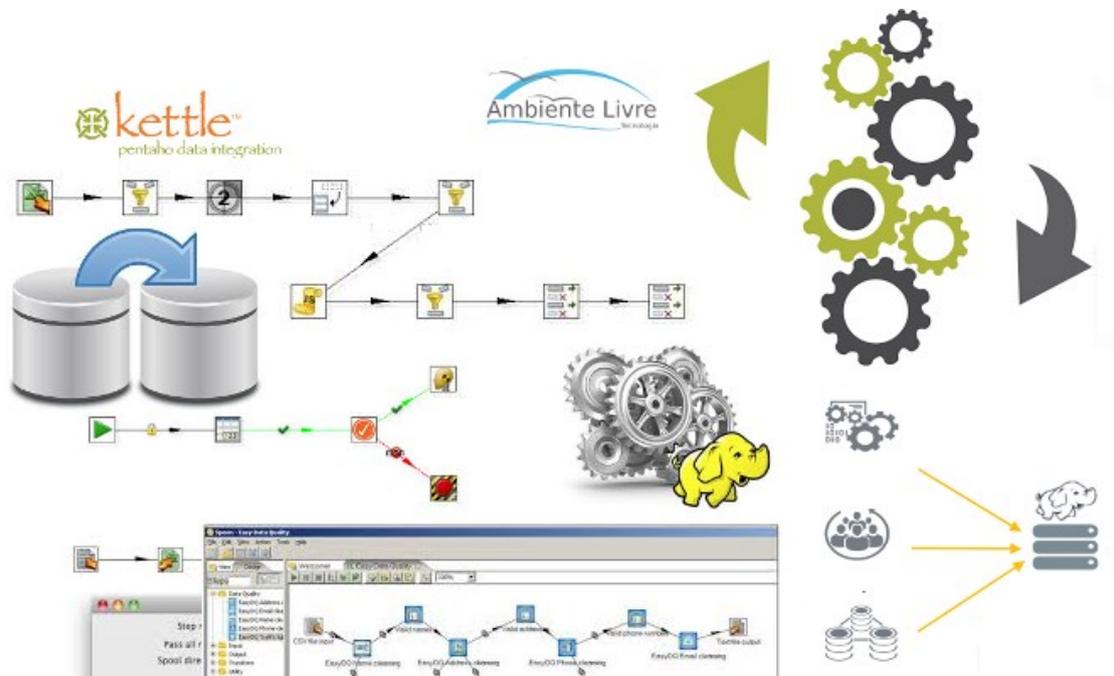
Depois, do nosso modelo dimensional lógico criado pelo Architect, foi iniciada a etapa de criar o modelo dimensional físico no Banco de Dados Relacional PostgreSQL. Para isso, primeiro fizemos a conexão do modelo no Architect com o referido Banco.

5.2.6 ETL

Com o modelo dimensional físico criado no PostgreSQL começou o processo de ETL, cuja definição consta do nosso capítulo 2.

Nosso ETL neste ponto, tratou basicamente da carga dos dados para dentro do modelo dimensional físico no PostgreSQL. Esta etapa, foi reduzida à carga, pois os dados na planilha do Excel estavam “limpos”, a extração e transformação foram realizadas na etapa de edição das planilhas no Excel.

Figura 19: Um desenho genérico de ETL.

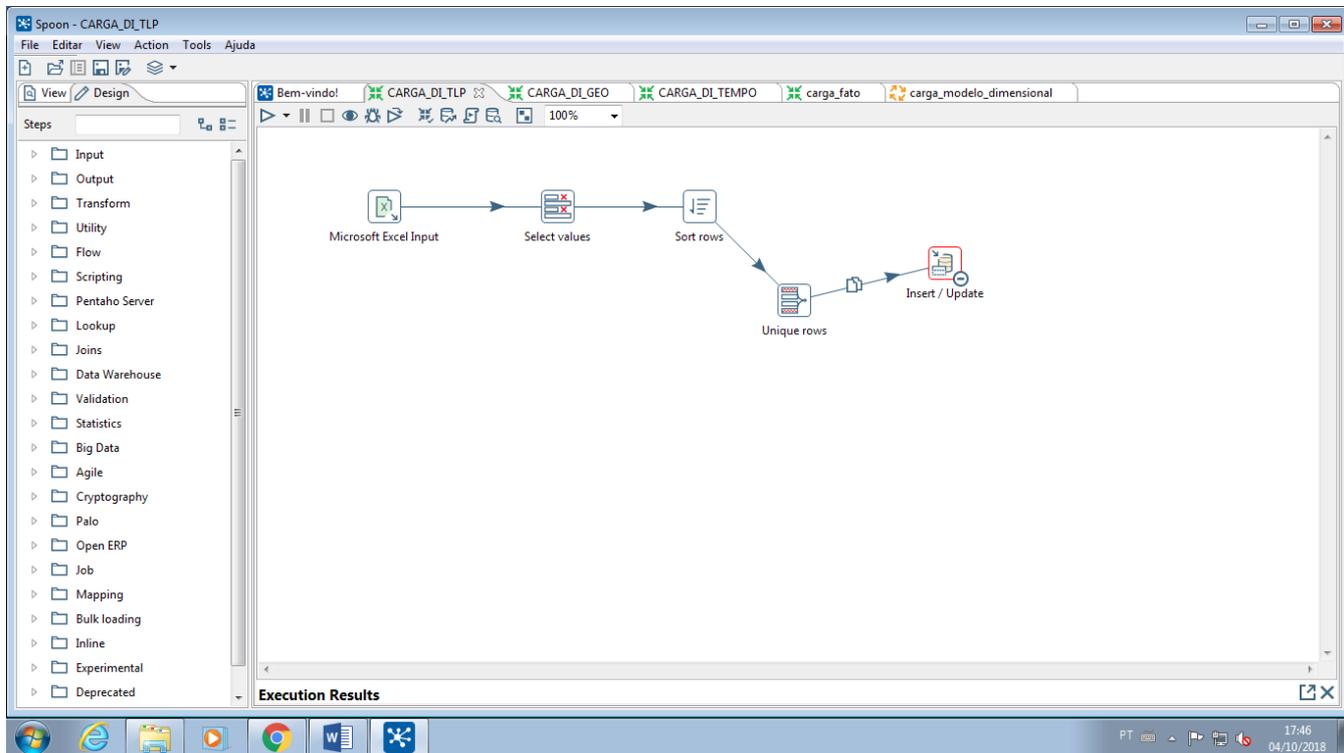


Fonte: www.ambientelivre.com.br/treinamento/pentaho-data-integration-kettle.html

5.2.6.1 A CARGA

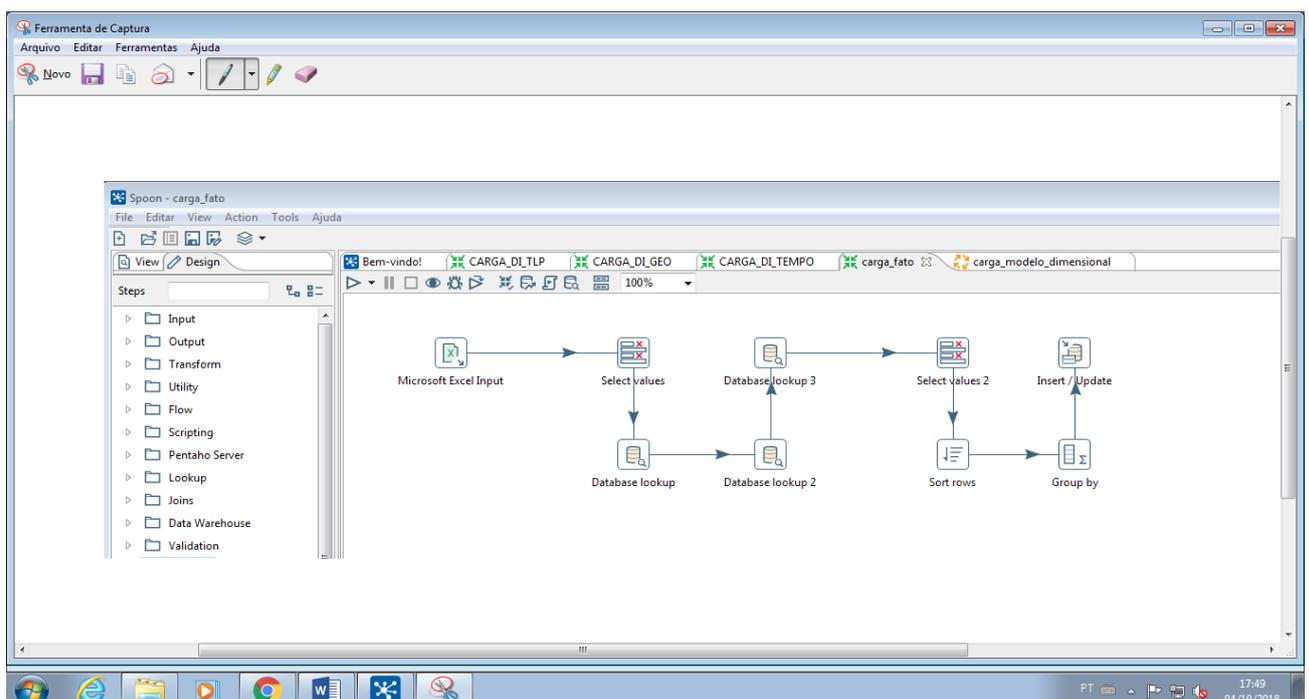
Para a etapa de carga, foi utilizado o Pentaho Integration (Kettle).

Figura 20: Carga da tabela DI_TLP. No primeiro “quadrado” (início da carga) os dados estão no Excel e no último “quadrado” eles já estão na tabela física do PostgreSQL.



Fonte: Autor (2018)

Figura 21: Carga da tabela fato Locação (FT_LOCACAO). No primeiro “quadrado” (início da carga) os dados estão no Excel e no último “quadrado” estão na tabela física do PostgreSQL.



Fonte: Autor (2018)

5.2.7 A VISUALIZAÇÃO E INFERÊNCIAS

Na etapa de visualização e inferências, foi utilizada a ferramenta Power BI da Microsoft. Primeiramente, foi selecionada a opção Obter Dados na aba Base.

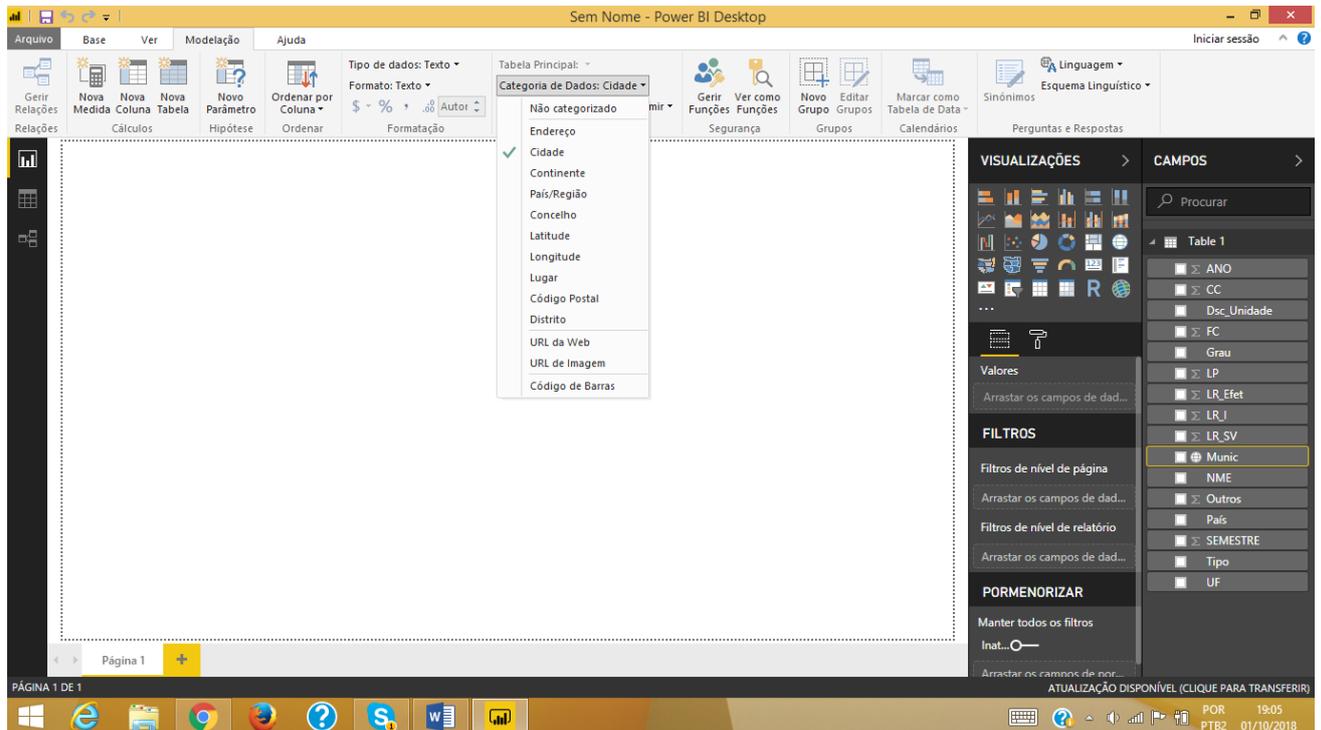
Figura 22: Obter dados na base.

Grau	Tipo	Disc_Unidade
1	outros	CGJ/JC - NUCLEO IV
2	outros	PRES.TJ - GABINETE DA PRESIDENCIA
2	outros	PRES.TJ - GABINETE DA PRESIDENCIA
0	outros	CGJ/JC - NUCLEO IV
2	outros	DCDP - ASSESSORIA DE CADASTRAMENTO
2	outros	DCDP - ASSESSORIA DE CADASTRAMENTO
0	outros	CGJ - ASSESSORIA CORREICIONAL
2	outros	GD - DES. JOSE ANTONIO TORRES MARQUES
2	outros	GD - DES. RICARDO OROFINO DA LUZ FONTES
2	outros	GD - DES. JOSE ANTONIO TORRES MARQUES
2	outros	GD - DES. RICARDO OROFINO DA LUZ FONTES
2	outros	1A. VICE-PRESIDENCIA
2	outros	1A. VICE-PRESIDENCIA
2	outros	GD - DES. ALEXANDRE DIVANENKO
2	outros	GD - DES. ALEXANDRE DIVANENKO
2	outros	GD - DES. RODOLFO CEZAR RIBEIRO DA SILVA TRIDAPALLI
Ambos	outros	GP - AUDITORIA INTERNA
2	outros	O2A. VICE-PRESIDENCIA
2	outros	O2A. VICE-PRESIDENCIA
1	outros	BLUMENAU - E.E. - GAB. JUIZ SUBSTITUTO
0	outros	PRES.TJ - AUDITORIA INTERNA
1	outros	BALCAMBORIU - E.E. - GAB. JUIZES SUBSTITUTOS

Fonte: Autor (2018)

Em seguida, foi necessário, modelar o campo Munic (município) e categorizando-o como Cidade. Faz-se isso, na aba Modelação, opção Categoria de Dados.

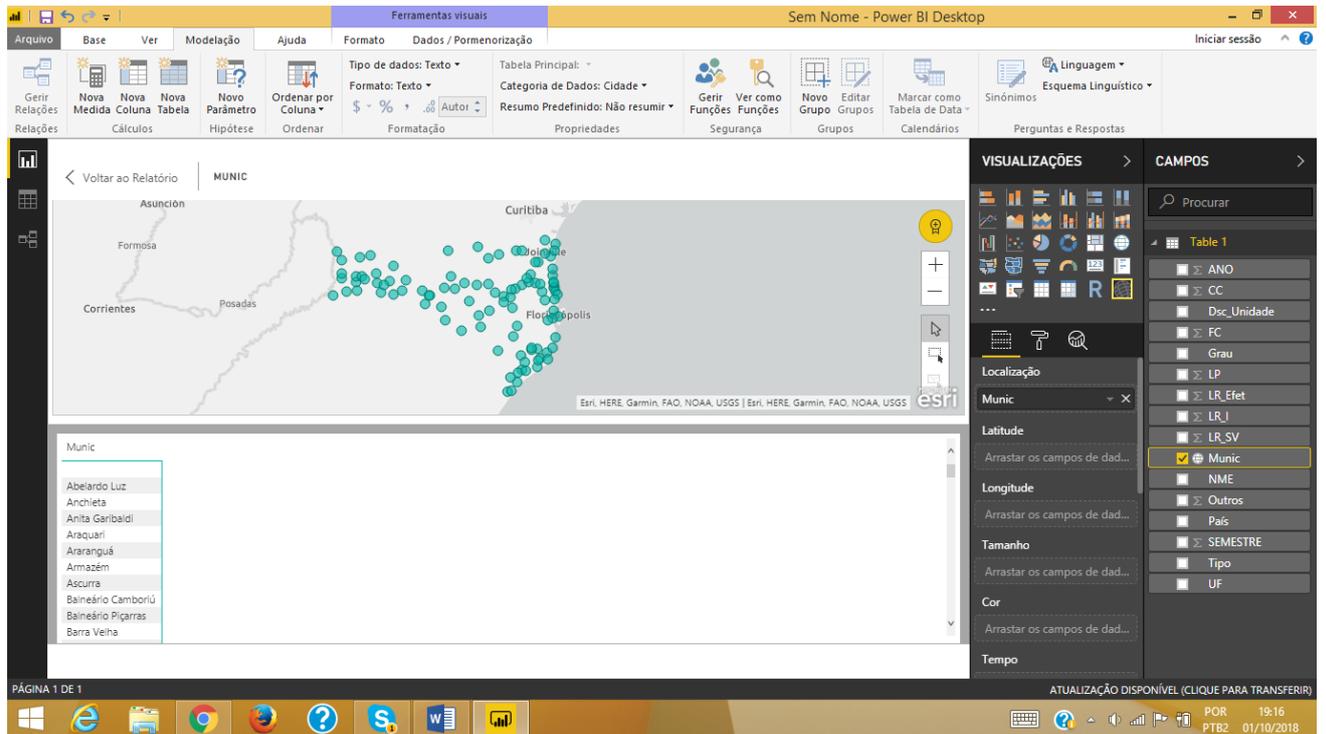
Figura 23: Dado Município já categorizado.



Fonte: Autor (2018)

A primeira, inferência, foi somente por municípios. Para isso, foi utilizado o mapa ArcGIS, que é um sistema de informação geográfica, utilizado para criar, gerenciar, compartilhar e analisar dados espaciais (o GIS quer dizer, *Geographic Information System*, em inglês), disponível e de fácil utilização no Power BI.

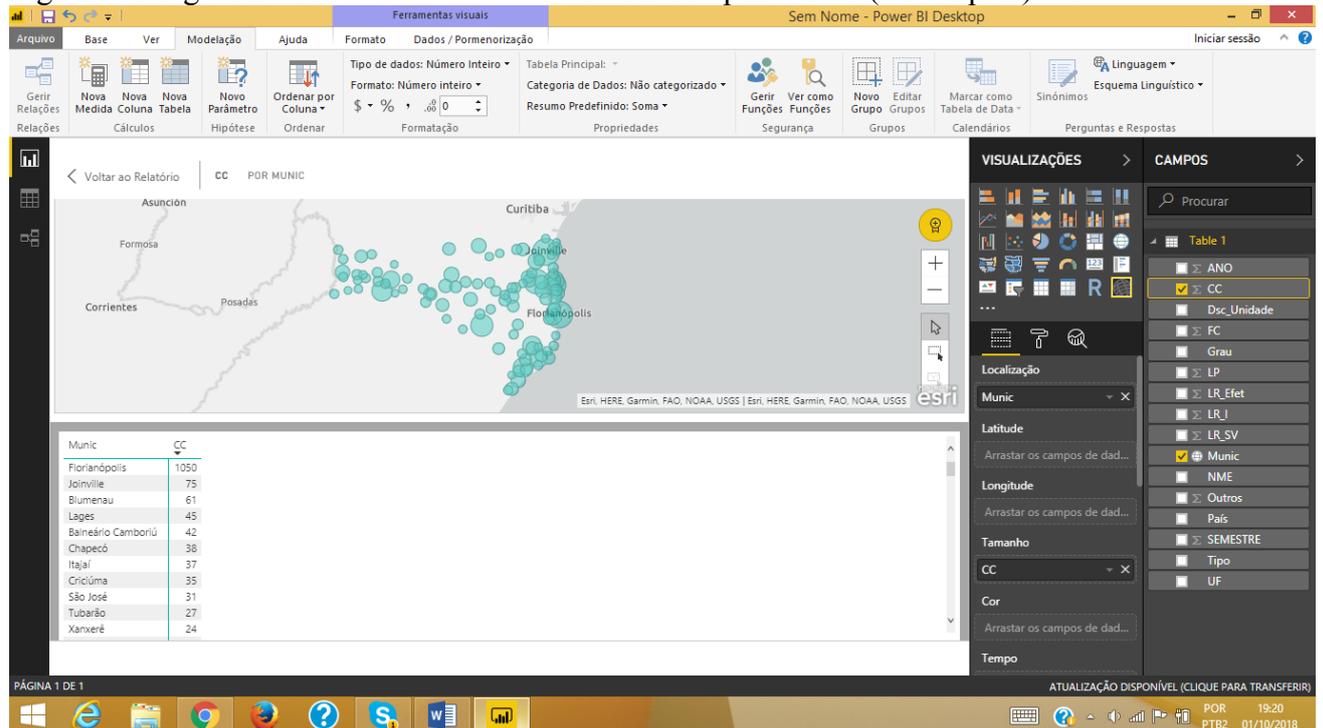
Figura 24: Primeira inferência Municípios no ArcGIS.



Fonte: Autor (2018)

A segunda inferência, “cruzou” os municípios, com o campo CC (Cargo em Comissão).

Figura 25: Segunda inferência – “cruzamento” dos campos Munic (municípios) e CC

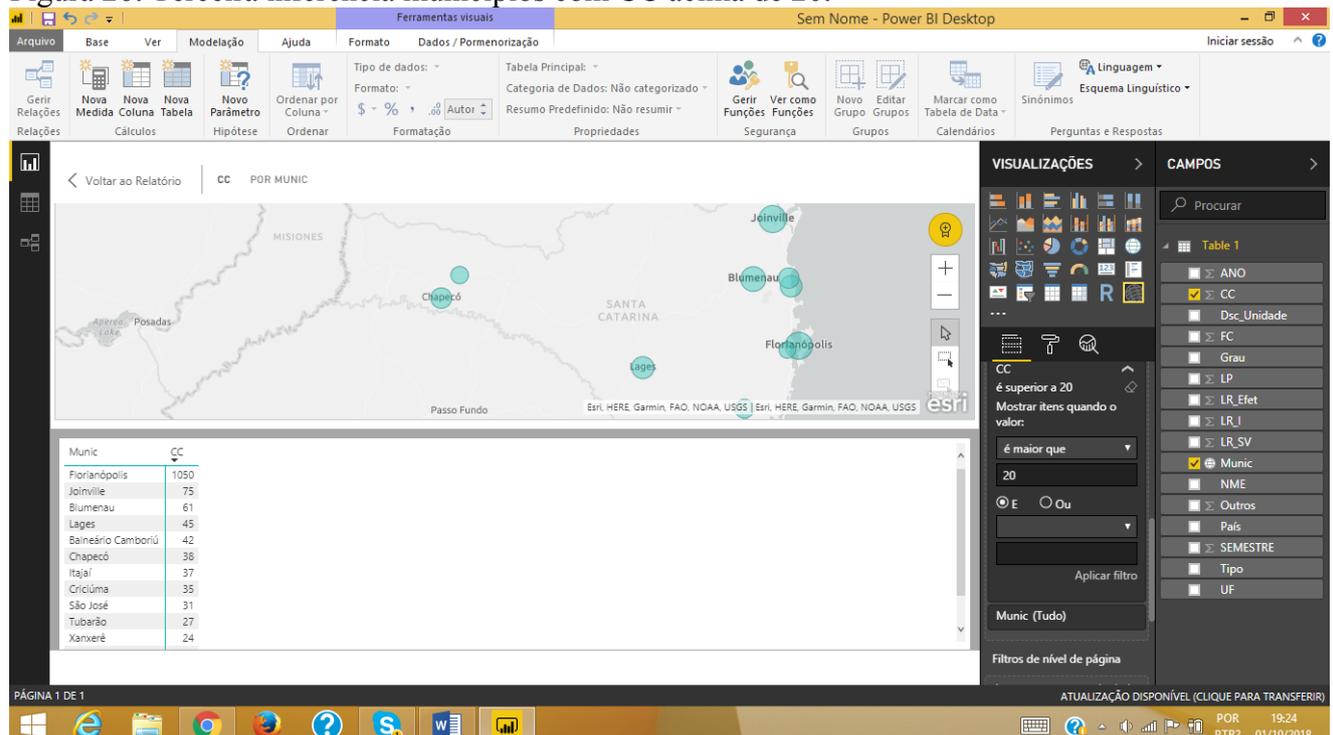


Fonte: Autor (2018)

Os círculos maiores, representam um número maior de cargos comissionados nestes municípios.

Na próxima tela, municípios com número acima de 20 cargos comissionados. Este gráfico em forma de mapa foi obtido, com o acréscimo de um filtro.

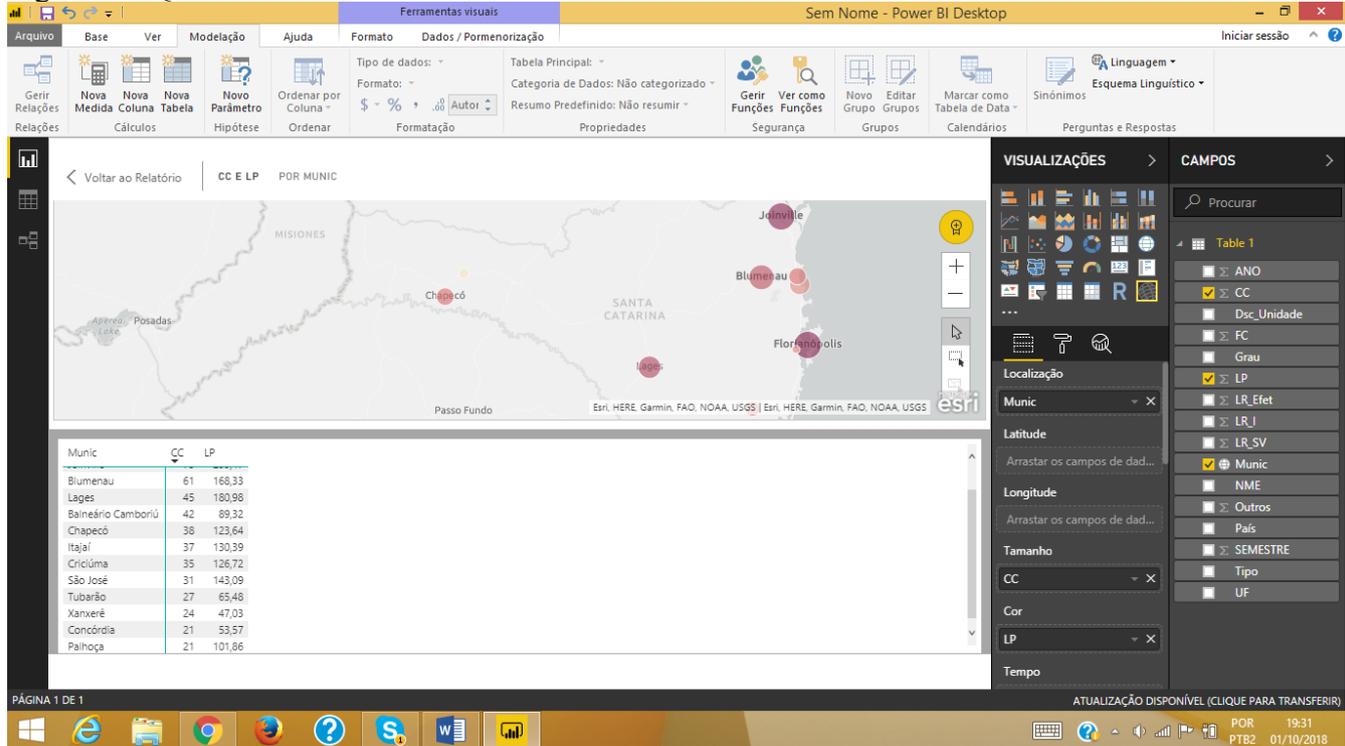
Figura 26: Terceira inferência municípios com CC acima de 20.



Fonte: Autor (2018)

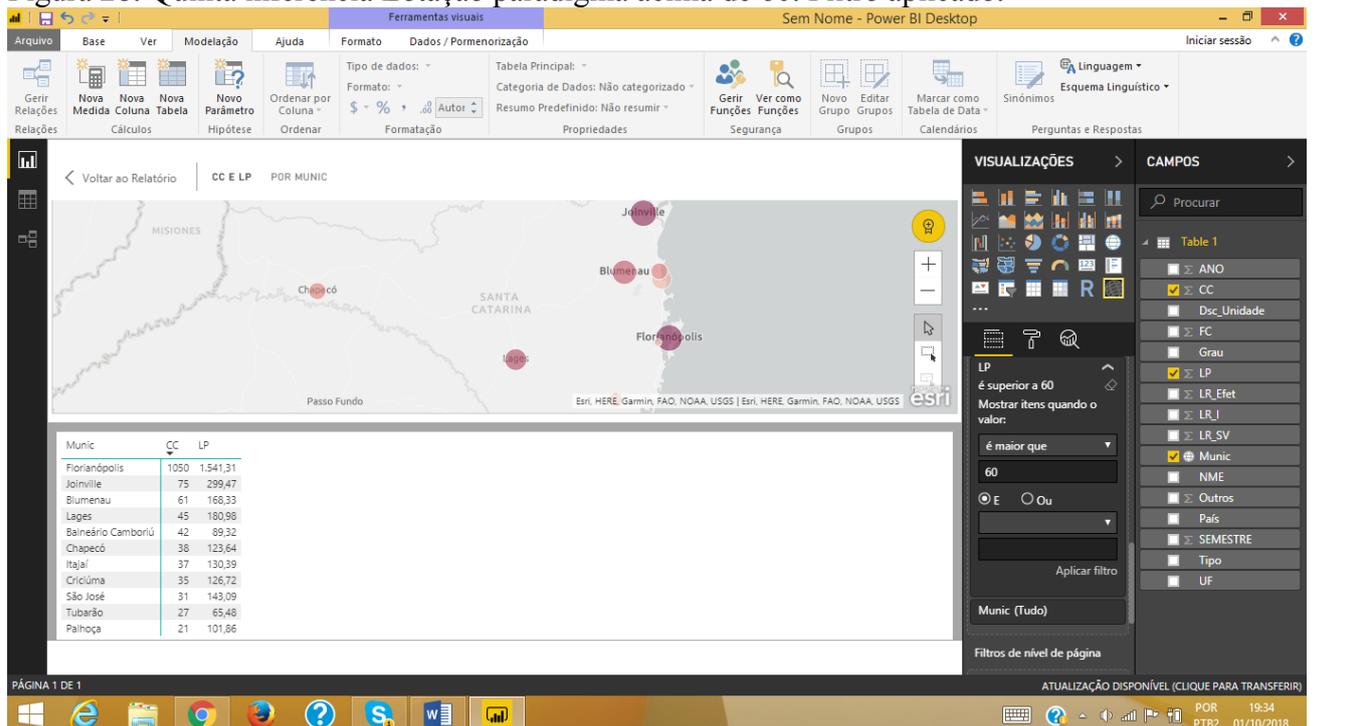
Na próxima imagem, foi feito um cruzamento entre CC e Lotação Paradigma (número sugerido pela Resolução 219/2016 do CNJ). A lotação paradigma é um número ideal para a lotação, conforme equação definida na citada resolução.

Figura 27: Quarta inferência CC acima de 20 com LP acima de zero.



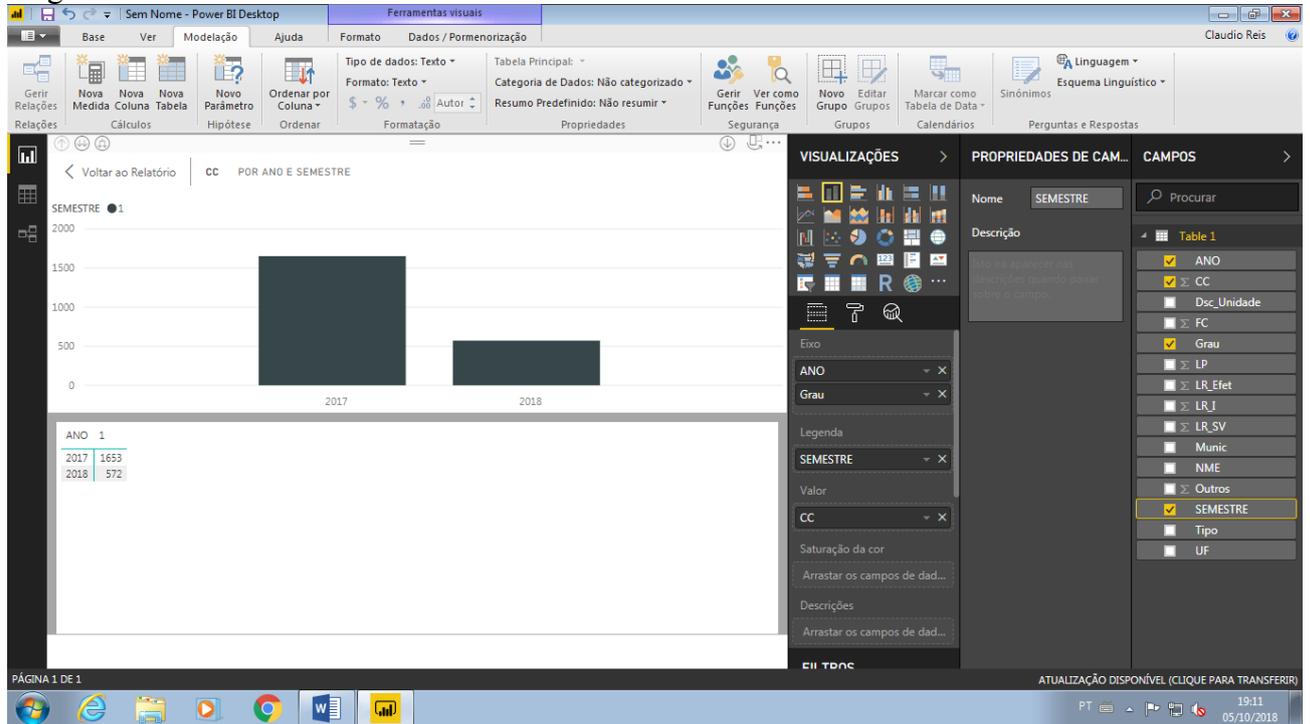
Fonte: Autor (2018)

Figura 28: Quinta inferência Lotação paradigma acima de 60. Filtro aplicado.



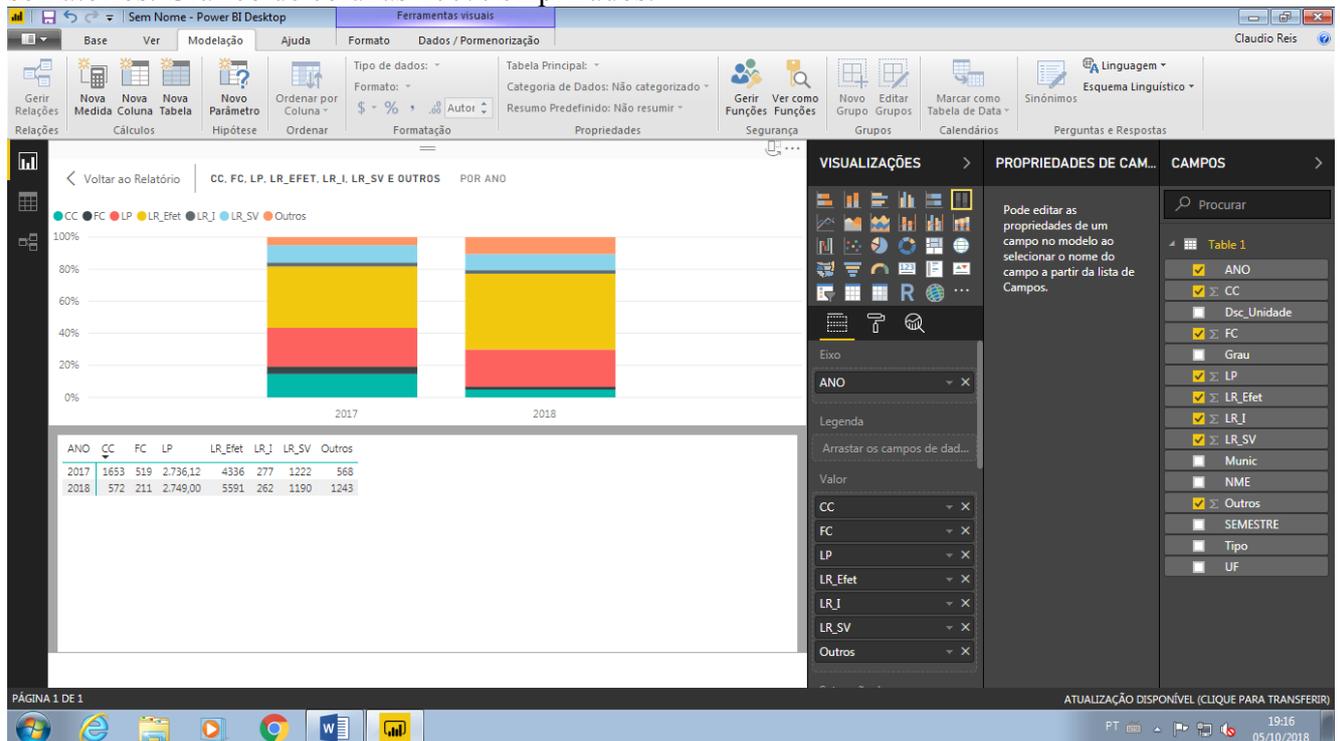
Fonte: Autor (2018)

Figura 29: Sexta inferência Gráfico de barras para Ano, Primeiro Semestre com número de cargos comissionados.



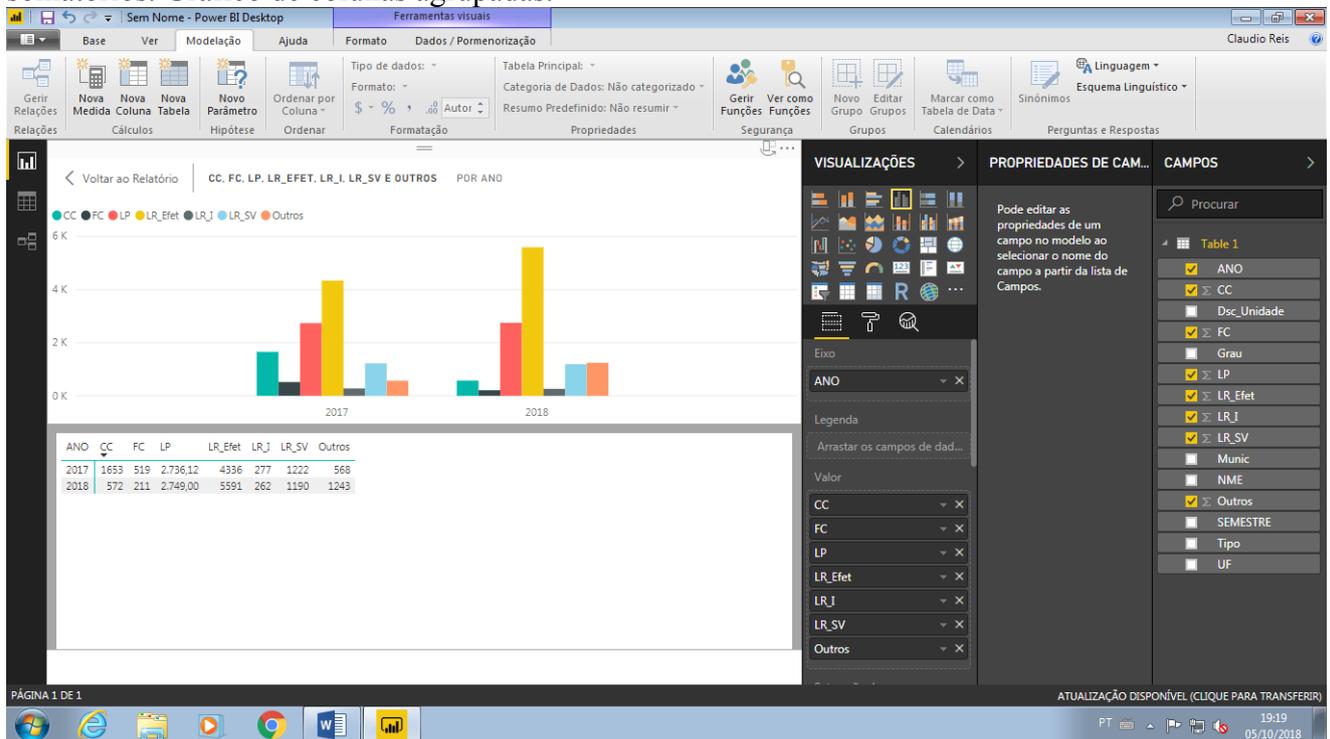
Fonte: Autor (2018)

Figura 30: Sétima inferência Ano, Primeiro Semestre com número de todos os campos somatórios. Gráfico de colunas 100% empilhados.



Fonte: Autor (2018)

Figura 31: Oitava inferência Ano, Primeiro Semestre com número de todos os campos somatórios. Gráfico de colunas agrupadas.



Fonte: Autor (2018)

5.3 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO

A aplicação da arquitetura de BI, permitiu a transformação de dados analíticos, planilhados inicialmente em arquivos PDF disponíveis no Portal da Transparência – Gestão de Pessoas – Lotação de Pessoal do Tribunal de Justiça de Santa Catarina, para um modelo dimensional que sintetizou em informações e se materializou visualmente em gráficos versáteis que permitiram diversos tipos de inferência.

6 CONCLUSÕES E TRABALHOS FUTUROS

Este capítulo tratará as conclusões deste projeto de *Business Intelligence*, bem como trabalhos futuros sobre o tema abordado.

6.1 CONCLUSÕES

O objetivo geral do projeto foi uma construção de arquitetura de *Business Intelligence*. Um dos itens desta arquitetura é o modelo dimensional, sendo este o primeiro objetivo específico. O modelo dimensional foi construído e foi persistido em um Sistema Gerenciador de Banco de Dados (SGBD).

O segundo objetivo específico, é a realização do processo de ETL. Os dados do Tribunal de Justiça de Santa Catarina foram carregados na base de dados. Este é o segundo elemento de uma arquitetura de BI.

O terceiro elemento da arquitetura de BI é a apresentação, como citado no terceiro objetivo específico. A área de visualização foi construída como mostrado no capítulo 5 (Desenvolvimento), do presente projeto.

Concluindo, este projeto facilitou a transparência e o acesso a informações que estavam dentro dos dados disponíveis nas tabelas (TLPs) em PDF, por meio de visualizações e interações, executando assim sua proposta inicial de transformar dados em informações que pudessem ser facilmente alcançadas pelo cidadão catarinense.

6.2 TRABALHOS FUTUROS

Com vistas à ampliação do projeto realizado (e também por limite de tempo) sugere-se para o futuro um estudo mais aprofundado da Resolução 219/2016 do CNJ que dá base às

tabelas de Lotação – Gestão de Pessoas Portal de Acesso à Informação e Transparência do Tribunal de Justiça de Santa Catarina e também:

- Com a finalidade de dar celeridade ao processo, sugere-se que as etapas extração e transformação dos dados das tabelas de Lotação – Gestão de Pessoas Portal de Acesso à Informação e Transparência do Tribunal de Justiça de Santa Catarina sejam realizadas de forma automatizada;

- Sugere-se colocar as visualizações e inferências em um site para pesquisa de qualquer usuário/cidadão;

- Sugere-se colocar o BI em um dispositivo móvel para a consulta a qualquer tempo e lugar;

- Sugere-se a expansão das visualizações e inferências para o número de funcionários total;

- Sugere-se expandir as inferências para produtividade, relacionando-se com o número de processos conclusos, população e a folha de pagamento do Poder Judiciário Catarinense;

- Sugere-se a criação da visualização de um modelo de crescimento vegetativo;

- Sugere-se expandir o projeto para outros tribunais;

- Sugere-se a criação de um modelo (paradigma) que sirva de referência entre tribunais.

Estes são os trabalhos futuros.

REFERÊNCIAS

BARBIERI, Carlos. BI – Business Intelligence: Modelagem e Tecnologia. Rio de Janeiro: Axcel Books, 2001.

Conselho Nacional de Justiça. Acesso à Informação. Portal CNJ, 2015. Disponível em: www.cnj.jus.br/transparencia/acesso-a-informacao-sic. Acesso em 29 de setembro de 2018.

Conselho Nacional de Justiça. Resolução n. 219, de 26 de abril de 2016. Distribuição de servidores, de cargos em comissão e de funções de confiança nos órgãos do Poder Judiciário de primeiro e segundo grau e outras providências, Brasília, DF, abril 2016. Disponível em: www.cnj.jus.br/busca-atos-adm?documentos=3110. Acesso em 29 de setembro de 2018.

INMON, William H. Como Construir o Data Warehouse. Tradução de Anamaria Netto Guz. 2. ed. Rio de Janeiro: Editora Campus, 1997.

KIMBALL, R.. Data Warehouse toolkit: O guia completo para modelagem dimensional. Rio de Janeiro, Editora Campus, 2002.

LAUDON, Kenneth C.; LAUDON, Jane Price. Sistemas de Informação. 4.ed. LTC: Rio de Janeiro, 1999.

MACHADO, Felipe Nery Rodrigues. Tecnologia e Projeto de Data Warehouse: uma visão multidimensional. São Paulo. Érica, 1a ed. 2004.

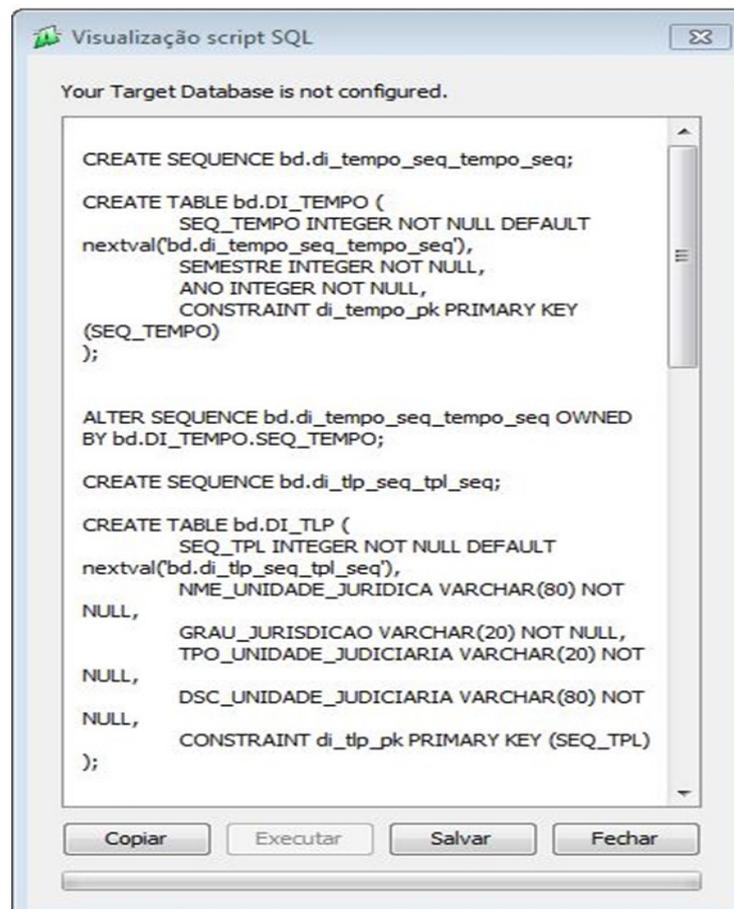
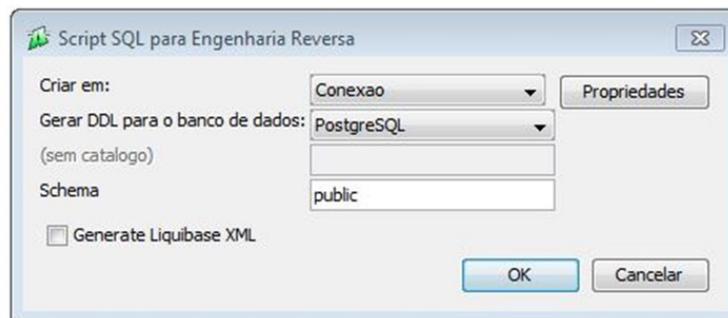
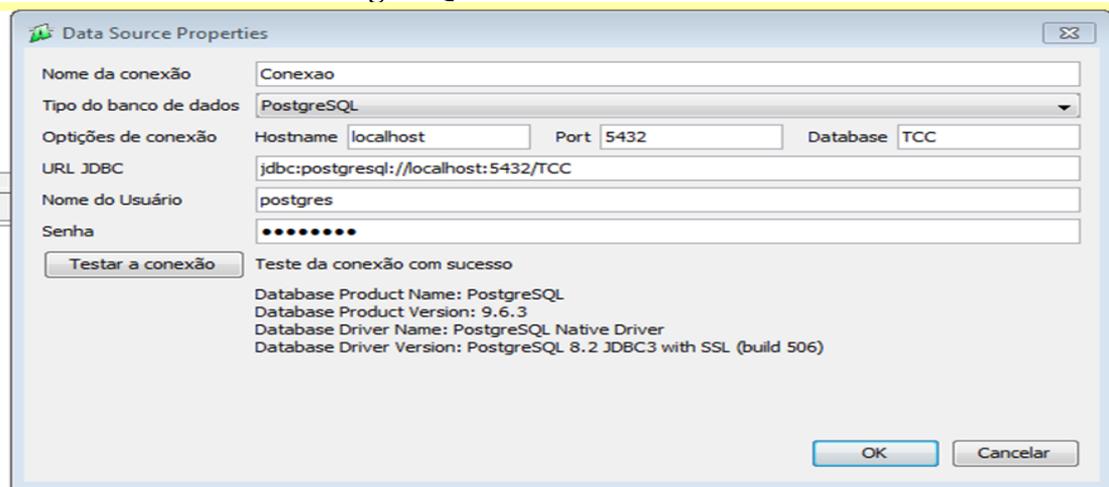
ROSINI, Alessandro Marco; PALMISANO, Angelo. Administração de sistemas de informação e a gestão do conhecimento. São Paulo, Thomson, 2003.

SILVA, Edna Lúcia da; MENEZES, Estera Muszkat. Metodologia da pesquisa e elaboração de dissertação. 3a Edição. Florianópolis: UFSC, 2001.

SOMMERVILLE, Ian. Engenharia de Software. Tradução: Selma Shin Shimizu Melnikoff, Reginaldo Arakaki, Edilson de Andrade Barbosa. 8ª. Edição São Paulo: Person Addison-Wesley, 2007.

ANEXOS

ANEXO A – Criação da Conexão SQL Power Architect scripts de conexão com o BD PostgreSQL e tabelas físicas criadas



Visualização script SQL

Your Target Database is not configured.

```

ALTER SEQUENCE bd.di_tlp_seq_tpl_seq OWNED BY
bd.DI_TLP.SEQ_TPL;

CREATE SEQUENCE bd.di_geografia_seq_geografia_seq;

CREATE TABLE bd.DI_GEOGRAFIA (
  SEQ_GEOGRAFIA INTEGER NOT NULL DEFAULT
nextval('bd.di_geografia_seq_geografia_seq'),
  NME_MUNICIPIO VARCHAR(80) NOT NULL,
  SGL_ESTADO VARCHAR(2) NOT NULL,
  CONSTRAINT di_geografia_pk PRIMARY KEY
(SEQ_GEOGRAFIA)
);

ALTER SEQUENCE bd.di_geografia_seq_geografia_seq
OWNED BY bd.DI_GEOGRAFIA.SEQ_GEOGRAFIA;

CREATE TABLE bd.FT_LOCACAO (
  SEQ_TEMPO INTEGER NOT NULL,
  SEQ_GEOGRAFIA INTEGER NOT NULL,
  SEQ_TPL INTEGER NOT NULL,
  TOTAL_LP DOUBLE PRECISION NOT NULL,
  TOTAL_LR_EFET INTEGER NOT NULL,
  TOTAL_LR_I INTEGER NOT NULL,
  TOTAL_LR_SV INTEGER NOT NULL,
  TOTAL_CC INTEGER NOT NULL,
  TOTAL_FC INTEGER NOT NULL,
  TOTAL_OUTROS INTEGER NOT NULL,
  CONSTRAINT ft_locacao_pk PRIMARY KEY
(SEQ_TEMPO, SEQ_GEOGRAFIA, SEQ_TPL)
);

```

Copiar Executar Salvar Fechar

Visualização script SQL

Your Target Database is not configured.

```

TOTAL_LR_SV INTEGER NOT NULL,
TOTAL_CC INTEGER NOT NULL,
TOTAL_FC INTEGER NOT NULL,
TOTAL_OUTROS INTEGER NOT NULL,
CONSTRAINT ft_locacao_pk PRIMARY KEY
(SEQ_TEMPO, SEQ_GEOGRAFIA, SEQ_TPL)
);

ALTER TABLE bd.FT_LOCACAO ADD CONSTRAINT
di_tempo_ft_locacao_fk
FOREIGN KEY (SEQ_TEMPO)
REFERENCES bd.DI_TEMPO (SEQ_TEMPO)
ON DELETE NO ACTION
ON UPDATE NO ACTION
NOT DEFERRABLE;

ALTER TABLE bd.FT_LOCACAO ADD CONSTRAINT
di_tlp_ft_locacao_fk
FOREIGN KEY (SEQ_TPL)
REFERENCES bd.DI_TLP (SEQ_TPL)
ON DELETE NO ACTION
ON UPDATE NO ACTION
NOT DEFERRABLE;

ALTER TABLE bd.FT_LOCACAO ADD CONSTRAINT
di_geografia_ft_locacao_fk
FOREIGN KEY (SEQ_GEOGRAFIA)
REFERENCES bd.DI_GEOGRAFIA (SEQ_GEOGRAFIA)
ON DELETE NO ACTION
ON UPDATE NO ACTION
NOT DEFERRABLE;

```

Copiar Executar Salvar Fechar

