

**UNIVERSIDADE DE CAXIAS DO SUL**  
**ÁREA DE CONHECIMENTO DE CIÊNCIAS EXATAS E ENGENHARIAS**

**LUCAS ELLWANGER**

**PROPOSTA DE UM *DATA WAREHOUSE* E VISUALIZAÇÃO DOS DADOS PARA O  
MÓDULO DE BI DO PORTAL LACOS WEB**

**CAXIAS DO SUL**

**2018**

**LUCAS ELLWANGER**

**PROPOSTA DE UM *DATA WAREHOUSE* E VISUALIZAÇÃO DOS DADOS PARA O  
MÓDULO DE BI DO PORTAL LACOS WEB**

Trabalho de Conclusão de Curso apresentado como  
requisito parcial para obtenção do grau de Bacharel em  
Ciência da Computação da Universidade de Caxias do Sul.  
Orientador: Prof. Dr. Daniel Luís Notari

**CAXIAS DO SUL**

**2018**

## RESUMO

O projeto LACOS (Lagoas Costeiras) tem por objetivo promover o estudo e disseminar o conhecimento sobre os recursos hídricos no litoral médio e sul do Rio Grande do Sul. Suas principais atividades são a realização de cursos, eventos e exposições temáticas, visando atingir a população com estes conhecimentos adquiridos. Para melhor organizar os dados coletados pelos pesquisadores do projeto, a UCS iniciou a construção do portal LACOS Web, o qual atualmente já possui uma série de funcionalidades, como georreferenciamento de lagoas, cadastros básicos, consultas, importação e exportação de dados. O presente trabalho teve o intuito de propor a este portal a construção de um *Data Warehouse* e formas de visualização das informações armazenadas. Também são apresentadas uma série de artefatos que compõem o *Business Intelligence*, elucidando a importância e função de cada componente deste sistema, o qual teve de ser compreendido e desenvolvido a fim de atender a real demanda dos pesquisadores do LACOS.

**Palavras-chave:** Lagoas costeiras, *Business Intelligence*, banco de dados, LACOS Web, portal, *Data Warehouse*

## LISTA DE QUADROS

Quadro 1: Evolução do BI.....	12
Quadro 2: Data Warehouse vs Banco de Dados tradicional.....	15
Quadro 3: Trabalhos relacionados.....	22
Quadro 4: Características visuais.....	30
Quadro 5: Requisitos funcionais do sistema.....	33
Quadro 6: Requisitos funcionais do usuário.....	34
Quadro 7: Requisitos não funcionais.....	36
Quadro 8: Tabelas analisadas.....	37
Quadro 9: Processo de ETL.....	40
Quadro 10: Exemplo de métricas.....	41
Quadro 11: Filtros do BI – Físico/Químico.....	49
Quadro 12: Filtros do BI – Fauna.....	54
Quadro 13: Classes de abundância.....	58

## LISTA DE FIGURAS

Figura 1: Arquitetura de BI.....	9
Figura 2: Arquitetura de BI em camadas.....	13
Figura 3: Processo de ETL.....	15
Figura 4: Data Warehouse e abordagens top-down e bottom-up.....	16
Figura 5: Tabelas fato e dimensões.....	17
Figura 6: Esquema de estrela.....	18
Figura 7: Esquema floco de neve.....	18
Figura 8: Processo de Data Mining.....	19
Figura 9: Camada de apresentação.....	21
Figura 10: Arquitetura de BI para o cenário de evolução de serviços.....	23
Figura 11: Modelo dimensional proposto para o DW.....	24
Figura 12: Dashboard que relaciona informações sobre custos e utilização dos serviços.....	24
Figura 13: Dashboard utilizando Google Charts.....	25
Figura 14: Gráfico de horas por mês.....	26
Figura 15: Gráfico de quantidade de horas por atividade.....	27
Figura 16: Gráfico de matrículas.....	28
Figura 17: Dashboard UNIPAMPA.....	28
Figura 18: Data Mart de desempenho modelado.....	29
Figura 19: Interface principal do Portal LACOS Web.....	31
Figura 20: Modelo ER do LACOS Web.....	32
Figura 21: Diagrama de caso de uso.....	34
Figura 22: Diagrama do caso de uso.....	35
Figura 23: Arquitetura de software.....	36
Figura 24: Modelo dimensional.....	39
Figura 25: Procedimento de ETL para o LACOS Web.....	40
Figura 26: Protótipo de saturação de O2 agrupado por trimestre.....	41
Figura 27: Protótipo de tela.....	42
Figura 28: Gráfico montado em excel pelos pesquisadores mostrando as medições das coletas das lagoas.....	43
Figura 29: Gráficos atuais utilizado pelos pesquisadores.....	44
Figura 30: Gráfico atual utilizado.....	44
Figura 31: Script de criação das tabela d_unidade_medida, d_variaveis e d_tipo_medicao.....	45
Figura 32: Script de criação da tabela fato f_valor_medicao.....	45
Figura 33: Procedure INSERE_DW.....	46
Figura 34: Formas de gráficos da biblioteca Google Charts.....	47
Figura 35: Filtros do BI – Físico/Químico.....	48
Figura 36: Gráfico físico/químico por tipo de local - Barras.....	50
Figura 37: Gráfico físico/químico – Linhas (Profundidade no eixo X).....	51
Figura 38: Gráfico físico/químico – Linhas (profundidade no eixo Y).....	52

Figura 39: Gráfico físico/químico - Barras.....	53
Figura 40: Filtro do BI - Fauna.....	54
Figura 41: Gráfico da presença de animais por local - Pizza.....	56
Figura 42: Gráfico das coletas de animais por local e data - Colunas.....	56
Figura 43: Gráfico da presença de animais por locais – Barras com número de indivíduos.....	57
Figura 44: Gráfico da presença de animais por locais – Barras com classe de abundância.....	59
Figura 45: BI - Administrativo.....	60

## LISTA DE ABREVIATURAS E SIGLAS

BI	<i>Business Intelligence</i>
CPD	Central de processamento de dados
CRM	<i>Customer Relationship Management</i>
DASD	<i>Direct Access Storage Device</i>
DM	<i>Data Mart</i>
DSS	<i>Decision Support System</i>
DW	<i>Data Warehouse</i>
ERP	<i>Enterprise Resource Planning</i>
ETL	<i>Extract, Transform, Load</i>
IET	<i>Índice de Estado Tráfico</i>
IQ	<i>Índice Químico</i>
IQA	<i>Índice de Qualidade da Água</i>
KDD	<i>Knowledge-Discovery in Databases</i>
KPI	<i>Key Performance Indicator</i>
LACOS	Lagoas Costeiras
OLAP	<i>Online Analytical Processing</i>
OLTP	<i>Online Transaction Processing</i>
PDF	<i>Portable Document Format</i>
PHP	<i>Hypertext Reprocessor</i>
SGBD	Sistema Gerenciador de Banco de Dados
SQL	<i>Structured Query Language</i>
SSIS	<i>SQL Server Integration Service</i>
TI	Tecnologia da Informação
UCS	Universidade de Caxias do Sul

## SUMÁRIO

1	INTRODUÇÃO.....	8
1.1	PROBLEMA DE PESQUISA.....	10
1.2	QUESTÃO DA PESQUISA.....	10
1.3	OBJETIVO.....	10
1.3.1	OBJETIVO GERAL.....	10
1.3.2	OBJETIVOS ESPECÍFICOS.....	10
1.4	ESTRUTURA DO TEXTO.....	10
2	BUSINESS INTELLIGENCE.....	12
2.1	DATA WAREHOUSE.....	14
2.1.1	Data Marts.....	16
2.2	MODELO DIMENSIONAL.....	17
2.2.1	Esquema estrela.....	17
2.2.2	Floco de neve.....	18
2.3	DATA MINING.....	18
2.4	APLICAÇÕES FRONT-END.....	20
2.5	TRABALHOS RELACIONADOS.....	22
2.6	CONSIDERAÇÕES FINAIS.....	29
3	PROPOSTA DE SOLUÇÃO.....	31
3.1	SITUAÇÃO ATUAL.....	31
3.2	REQUISITO NECESSÁRIOS.....	33
3.2.1	Requisitos funcionais do sistema.....	33
3.2.2	Requisitos funcionais do usuário.....	34
3.2.3	Requisitos não funcionais.....	36
3.2.4	Data warehouse.....	36
3.2.5	INTERFACE GRÁFICA.....	41
4	IMPLEMENTAÇÃO.....	43
4.1	SITUAÇÃO ATUAL.....	43
4.2	CONSTRUÇÃO DO DATA WAREHOUSE.....	44
4.3	CAMADA DE APRESENTAÇÃO.....	46
4.4	CONSIDERAÇÕES FINAIS.....	47
5	ESTUDO DE CASO.....	48
5.1	MÓDULO DE BI.....	48
5.1.1	BI – FÍSICO/QUÍMICO.....	48
5.1.1.1	Gráfico físico/químico contendo valores médios de coletas por tipos de local.....	49
5.1.1.2	Gráfico biológico contendo valores médios de coletas por profundidade do local – Linhas (Profundidade no eixo X).....	50
5.1.1.3	Gráfico físico/químico contendo valores médios de coletas por profundidade do local – Linhas (Profundidade no eixo Y).....	51



5.1.1.4Gráfico físico/químico contendo valores médios de coletas por profundidade do local – Barras	52
5.1.2 BI – FAUNA	53
5.1.2.1Gráfico da presença de animais por local	55
5.1.2.2Gráfico da presença de espécies por locais	57
5.1.3 BI – ADMINISTRATIVO	59
5.2 QUESTIONÁRIO	60
6 CONCLUSÃO	62
6.1 TRABALHOS FUTUROS	63
7 REFERÊNCIAS BIBLIOGRÁFICAS	64
ANEXO A – EXEMPLO DE IMPLEMENTAÇÃO DE GRÁFICOS UTILIZANDO GOOGLE CHARTS	67

## 1 INTRODUÇÃO

A Planície Costeira do Rio Grande do Sul compreende uma vasta área do território leste gaúcho, estendendo-se desde a cidade do Chuí até Torres. Ao longo desta superfície encontram-se a Lagoa Mirim, Laguna dos Patos, Lagoa Mangueira e uma grande quantidade de lagoas menores, interligadas ou isoladas (SHÄFER, 2009).

O Projeto Lagoas Costeiras abrange a área dos municípios de Mostardas, Tavares, São José do Norte e Santa Vitória do Palmar as quais possuem recursos hídricos únicos devido a sua localização entre a Laguna dos Patos/Lagoa Mirim e o Oceano Atlântico. A partir dos anos 80 iniciou-se um estudo ecológico a cerca das lagoas costeiras do Rio Grande do Sul, onde foram evidenciadas as peculiaridades destas (SHÄFER, 2009).

Os dados dessas lagoas foram armazenadas em um banco de dados, e podem ser explorados por consultas pontuais ou através de uma análise mais sofisticadas, denominada *Business Intelligence*, ou simplesmente BI, o qual é um termo patenteado pela empresa Gartner. Apesar de parecer um termo moderno, seu conceito prático já data de sociedades antigas do Oriente Médio quando cruzavam informações obtidas da natureza em benefícios de suas aldeias. Passado décadas de evolução o conceito de BI segue inalterado (PRIMAK, 2008). Atualmente existem técnicas, tecnologias e softwares adequados para BI, porém BI não é algo tangível. É um processo que envolve métodos e técnicas de trabalho, recursos humanos, fontes de informação, entre outros fatores (LOH, 2014).

De acordo com Herrera (2003), um grande problema encontrado pelas empresas é a enorme quantidade de informações acumuladas em suas bases de dados, escondendo informações realmente importantes. Neste contexto, o objetivo do BI estaria em definir regras e técnicas para sintetizar esse grande volume de dados em um depósito estruturado de informações relevantes.

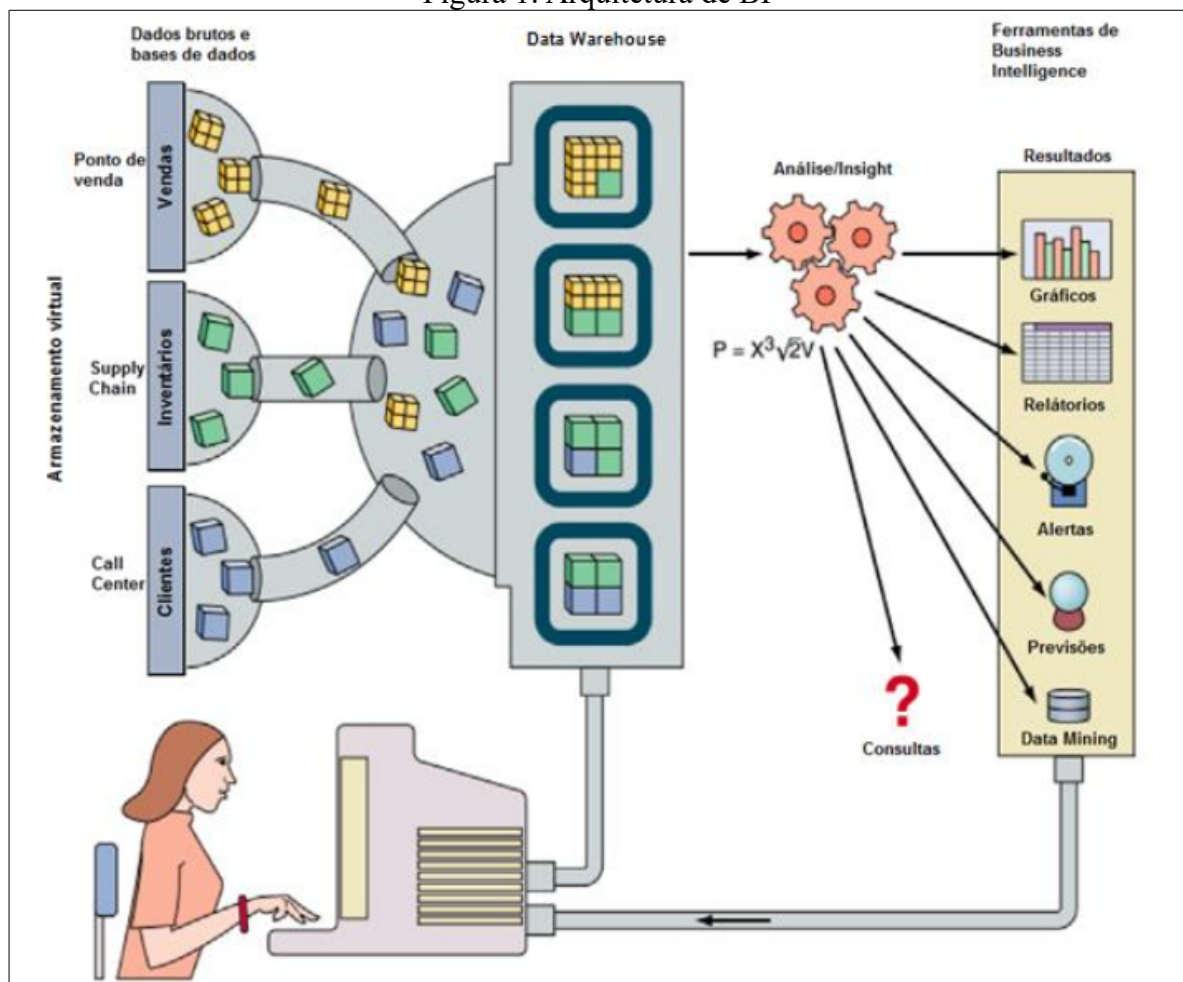
Segundo Silva (2011, p.32), *Business Intelligence* “consiste na transformação metódica e consciente dos dados provenientes de quaisquer fontes de dados (estruturados e não estruturados) em novas formas de proporcionar informação e conhecimento dirigidos aos negócios e orientados aos resultados”.

Para Aronson (et al., 2009), os principais objetivos do BI são prover acesso aos dados propiciando a capacidade rápida de identificação e análise destas informações no auxílio à tomadas de decisões com base em evidências sólidas que foram coletadas. Ou seja, o BI coleta e organiza os dados, processa de forma que se resumam apenas à informações relevantes e prevê possíveis cenários que deverão ajudar na tomada de decisões.

Um grande desafio ao se implantar um sistema para BI é a construção de uma ferramenta de consulta que consiga agregar em um único painel toda a informação considerada essencial, de forma imediata, sem muita complexidade e da forma mais atrativa possível. Essa apresentação pode ocorrer através de gráficos, relatórios, *dashboards*, tabelas, entre outras (COSTA, 2012). Ainda conforme Costa, para se alcançar o sucesso nessa tarefa é essencial que os dados já tenham sido coletados, classificados e armazenados adequadamente em um *Data Warehouse* (DW) para assim facilitar o acesso somente à informação relevante.

A Figura 1 exemplifica a arquitetura de um BI aplicado a um sistema de vendas. É possível ver que existem 3 fontes de dados brutos independentes: Ponto de venda, *Supply Chain*<sup>1</sup> e *Call Center*<sup>2</sup>. O segundo estágio é um *Data Warehouse*, o qual é alimentado pelas fontes de dados da primeira camada e que será consultado pelas ferramentas de visualização através de relatórios, gráficos, *dashboards*, entre outras formas.

Figura 1: Arquitetura de BI



Fonte: RANJAN (2009)

- 1 *Supply Chain*: é uma expressão inglesa que significa “cadeia de suprimentos” ou “cadeia logística”.
- 2 *Call Center*: central de atendimento que tem como objetivo fazer a interface entre o cliente e a empresa.

## 1.1 PROBLEMA DE PESQUISA

O projeto Lagoas Costeiras em parceria com a Universidade de Caxias do Sul e com pesquisadores das áreas de biologia e tecnologia da informação da UCS desenvolveram o Portal LACOS Web, centralizando e organizando as informações em apenas uma plataforma. Informações que até então estavam distribuídas em diferentes meios, tais como planilhas eletrônicas, formulários e pequenos bancos de dados.

Com todas as informações do projeto LACOS reunidas em apenas um banco de dados, identificou-se a possibilidade de visualizar diversas informações que este sistema possui, sumarizadas e exibidas em consultas pontuais. Dando assim aos pesquisadores a possibilidade de identificar rapidamente através de uma só consulta diversas informações essenciais sintetizadas.

## 1.2 QUESTÃO DA PESQUISA

Um sistema para BI consegue prover uma visão sintética dos dados do LACOS Web?

## 1.3 OBJETIVO

### 1.3.1 OBJETIVO GERAL

O objetivo geral do trabalho é construir um *Data Warehouse* e utilizar recursos de visualização dos dados para o portal LACOS Web.

### 1.3.2 OBJETIVOS ESPECÍFICOS

Para alcançar o objetivo geral deste trabalho, será necessário atingir os objetivos específicos elencados abaixo:

1. Elaborar um levantamento bibliográfico baseado em trabalhos relacionados.
2. Elaborar uma proposta de solução.
3. Implementar a proposta de solução.
4. Validar a integridade das informações com os usuários do LACOS Web.

## 1.4 ESTRUTURA DO TEXTO

Este trabalho está estruturado em 7 capítulos conforme descrito a seguir.

O capítulo 2 aprofunda-se a cerca de *Business Intelligence*, apresentando sua evolução, conceito e arquitetura. Também aborda sobre *Data Warehouse*, explanando sobre sua construção, diferenças que possui de um banco de dados tradicional, suas subdivisões em *Data Marts* e a forma de como construir um modelo dimensional. Também aborda sobre o processo de *Data Mining*, explicando a importância e como utilizar esta técnica. Neste

trabalho ainda, é apresentado componentes de aplicações *front-end*, e como são fundamentais para um BI de sucesso. Além de demonstrar através de trabalhos relacionados uma série de características peculiares de cada um dos sistemas de BI.

O capítulo 3 é focado na proposta de solução, momento onde são elencados os requisitos de sistema, elaborado protótipos da aplicação e construído o modelo dimensional para o BI. O capítulo 4 apresenta a implementação deste sistema proposto, retratando a situação atual, e evidenciando cada artefato utilizado para o desenvolvimento deste novo módulo, listando também as soluções e problemas encontrados. No capítulo 5 está dissertado sobre o que foi construído, apresentando todas as funcionalidades deste BI e apresentando cada forma de visualização possível das informações do *Data Warehouse*. O capítulo 6 apresenta as conclusões a respeito do que foi desenvolvido e propõe trabalhos futuros.

## 2 BUSINESS INTELLIGENCE

*Business Intelligence*, apesar de ser um termo recente, é um conceito prático e já era utilizado por povos antigos. Na medida que seu emprego propicia realizar várias análises, projeções, entendimento de processos e auxílio à tomada de decisões, o interesse pelo BI vem crescendo cada vez mais. E com a evolução natural da tecnologia, o BI ganhou cada vez mais abrangência dentro das empresas (PRIMAK, 2008). O Quadro 1, apresenta a evolução do BI:

Quadro 1: Evolução do BI

Período	O que aconteceu?
<b>Idade antiga</b>	– Povos do Oriente Médio cruzavam informações da natureza para analisarem a viabilidade de cultivo de certas espécies, períodos de pesca, entre outras ações.
<b>Século XVI</b>	– Rainha Elizabeth I, determinou que a base da força inglesa fosse “informação e comércio” e ordenou ao filósofo Francis Bacon que inventasse um sistema dinâmico de informação, o qual foi fielmente difundido pelos ingleses.
<b>Década de 60</b>	– Evolução e redução do tamanho dos computadores. – Empresas passaram a perceber os dados como uma possível e importante fonte geradora de informações decisórias e que renderiam eventuais lucros. – Época dos cartões perfurados e linguagens de programação: <i>COBOL</i> , <i>ASSEMBLY</i> e <i>FORTRAN</i> .
<b>Década de 70</b>	– Evolução do SGBD (Sistema Gerenciador de Banco de Dados) e DASD ( <i>Direct Access Storage Device</i> – Dispositivo de Armazenamento de Acesso Direto).
<b>Década de 80</b>	– Desenvolvimento e evolução das linguagens <i>CLIPPER</i> e <i>PASCAL</i> . – Início da aplicação do termo <i>Business Intelligence</i> .
<b>Década de 90</b>	– Início do uso termo CPD (Centro de Processamento de Dados) nas empresas. – Desenvolvimento do <i>Data Warehouse</i> . – Interesse pelas soluções de BI.
<b>Década de 2000</b>	– Evolução dos conceitos de DSS ( <i>Decision Support System</i> – sistema de suporte a decisão), Planilhas Eletrônicas, Geradores de Consulta e de Relatórios, <i>Data Marts</i> (DM), <i>Data Mining</i> , Ferramentas OLAP <sup>3</sup> , entre outras. – Estreitamento das conexões entre o BI e o ERP ( <i>Enterprise Resource Planning</i> ). – Surgimento do termo <i>e-business</i> <sup>4</sup> .

Fonte: PRIMAK (2008)

3 OLAP (*Online Analytical Processing*): capacidade de manipular e analisar um grande volume de dados sob múltiplas perspectivas.

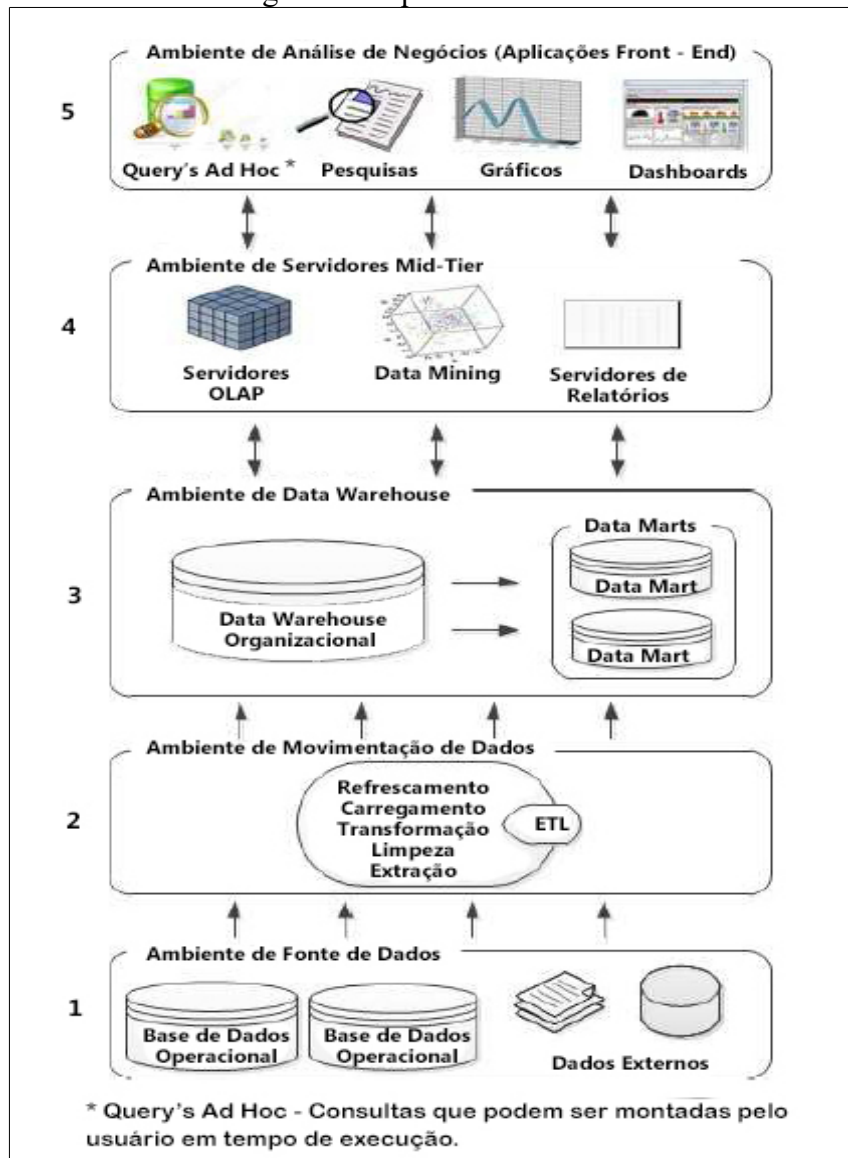
4 *E-business* (*Electronic Business*): termo utilizado para identificar negócios efetuados por meios eletrônicos.

A área de BI vem sendo considerada essencial dentro de qualquer empresa devido à quantidade e qualidade da informação que este processo fornece (TURBAN, SHARDA, DELEN, 2010), dando aos gestores uma visão ampla sobre as questões estratégicas, táticas e operacionais.

Segundo Santos e Ramos (2009), pode-se dizer que um sistema para BI é composto pelo processo de coleta de dados operacionais (podendo ser de fontes diversas), armazenamento em repositório adequado e posterior exploração destes dados de forma a gerar uma apresentação das informações essenciais emitidos por relatório, *dashboard*, gráficos, entre outras formas.

Na Figura 2 é ilustrado uma arquitetura geral de um sistema de BI, definido através de camadas.

Figura 2: Arquitetura de BI em camadas



Fonte: Adaptado de COSTA (2012)

Segundo Costa (2012), esta arquitetura é composta por cinco camadas:

**1. Ambiente de fonte de dados:** corresponde a todas as fontes de dados que irão compor o sistema, podendo ser de origem interna ou externa a organização, alguns exemplos podem ser: base de dados do ERP, folhas de cálculo e planilhas eletrônicas.

**2. Ambiente de movimentação de dados:** momento onde se realiza toda parte de ETL.

**3. Ambiente do *Data Warehouse*:** integra o *Data Warehouse* e seus *Data Marts*.

**4. Ambiente de servidor *Mid-Tier*:** Camada onde são trabalhados os dados vindo do *Data Warehouse*, através de recursos e técnicas para sintetização dos dados a somente informações essenciais. Nesta camada são utilizadas técnicas de OLAP e *Data Mining*.

**5. Ambiente de análise de negócio (aplicações *front-end*):** Parte da aplicação que interage com o usuário, a qual permite o acesso e manuseio da informação, tais como gráficos, relatórios e *dashboards*.

Conforme Amarante (2014), pode-se identificar os 3 primeiros níveis como entrada de dados, também denominada por alguns autores como *Data Warehousing*, pois é o momento em que os dados são extraídos de suas fontes, transformados e alocados no armazém de dados.

Os outros dois níveis, apresentam a saída de dados, que dará aos usuários a capacidade de conhecimento e respostas a questões específicas, providas através de alguma representação visual (AMARANTE, 2014). Um sistema para BI não é composto somente por *Data Warehouse*, é imprescindível que a apresentação gráfica dos dados tenham um aspecto agradável e boa usabilidade (FRANCISCO, 2009).

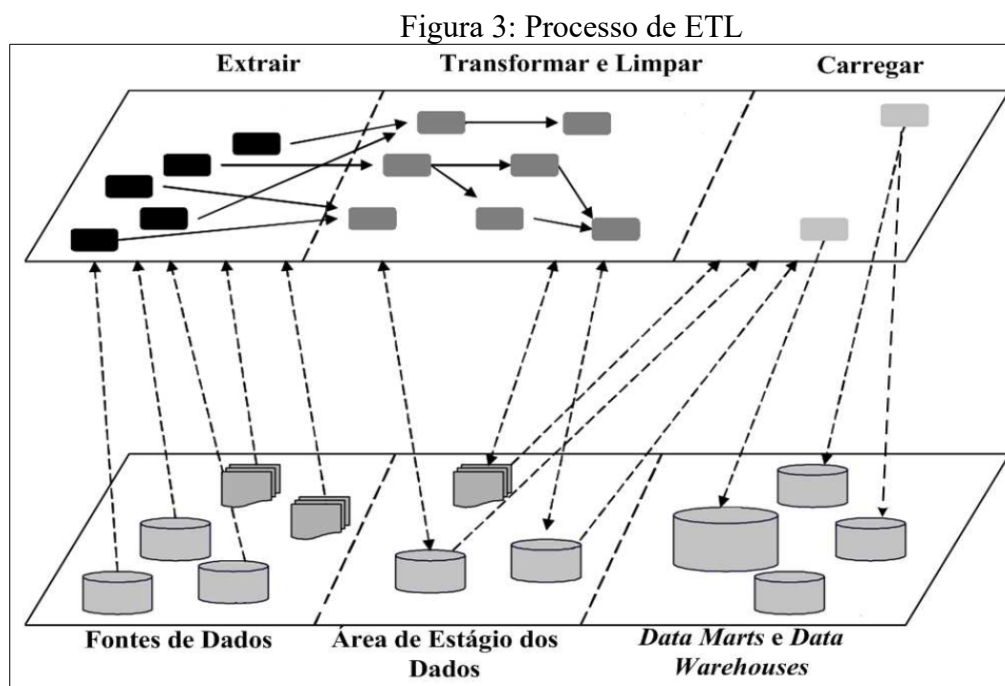
De acordo com Francisco (2009), sistemas de *Business Intelligence* contribuem para a melhora a inteligência cognitiva, aumentam a capacidade de aprendizagem da organização e melhoram a criatividade organizacional.

## 2.1 DATA WAREHOUSE

Um importante passo para o sucesso do BI é a construção do *Data Warehouse*, o qual é um armazém de dados corporativos integrado, único, centralizado e seguro, alimentado por todas as fontes de dados da empresa. Uma importante etapa para implantação de um DW é o processo de Extração, Tratamento e Limpeza dos dados – ETL, pois é a fase onde os dados são extraídos de suas origens, padronizados, transformados e passam a integrar o ambiente do DW. Através do ETL é possível identificar redundâncias ou inconsistências de dados operacionais, ganhando também produtividade, pois serão armazenados somente dados



essenciais já processados (UNO, 2015).



Fonte: COSTA (2012)

Segundo Inmon (1999), um *Data Warehouse* deve ser “não-volátil”, ou seja, os dados inseridos não devem ser apagados, garantindo assim que consultas subsequentes gerem o mesmo resultado, sendo capaz assim de armazenar dados históricos e atuais. Desta maneira possibilita comparações, identificações de tendências, previsões de cenários, entre outras possibilidades. Outra característica de um DW é que os dados são carregados de forma esporádica, a fim de não colocar em risco a rapidez e a precisão das informações.

Existem uma série de diferenças entre um *Data Warehouse* e um banco de dados tradicional, as quais são exibidas pelo Quadro 2:

Quadro 2: *Data Warehouse* vs Banco de Dados tradicional

<b><i>Data Warehouse</i></b>	<b>Banco de dados tradicional</b>
Atividades baseadas em análises orientadas à decisão	Atividades operacionais
Menos frequentes	Mais frequentes
Menos previsíveis	Mais previsíveis
Maiores quantidade de dados por consulta	Menos quantidades de dados acessados por consulta
Consultas principalmente de dados derivados	Consulta principalmente de dados primitivos
Exige dados passados, presentes e projetados	Exige principalmente dados atuais
Muitas derivações complexas	Pouca ou nenhuma derivação complexa

Fonte: SANTOS (2006)

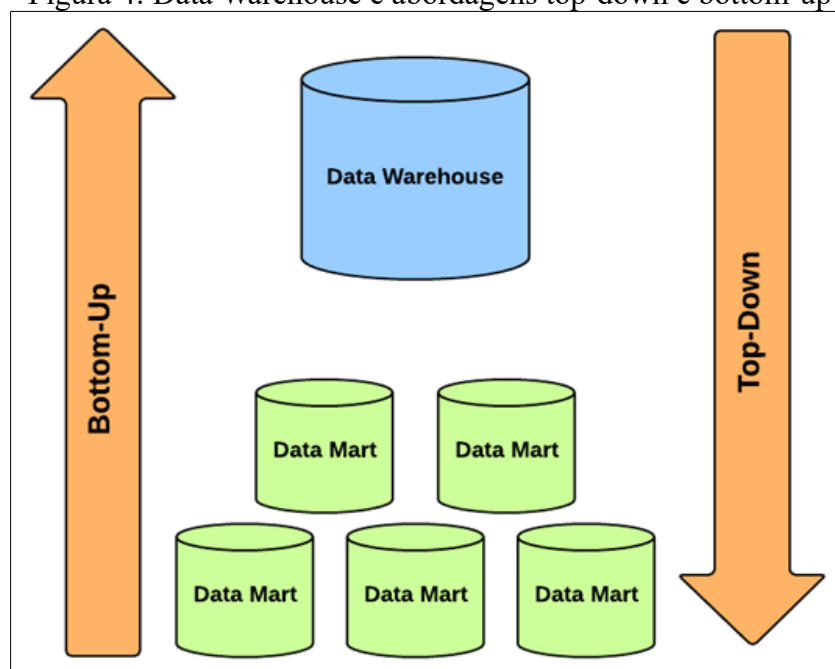
### 2.1.1 Data Marts

Conforme Elias (2014), *Data Marts* são as divisões do DW em subconjuntos separados por conteúdos específicos. Inmon (1999) descreve *Data Marts*, como estruturas de dados que contêm informações de acordo com algum interesse e necessidade, podendo assim restringir o acesso aos dados, ganhando eficiência. O autor ainda é favor de que o *Data Warehouse* seja derivado de diversos *Data Marts*, organizados e orientados por assuntos chave, de acordo com as necessidades específicas dos utilizadores finais do sistema, mantendo somente armazenados os dados relevantes aos processos específicos. Definição a qual defronta com Kimball(2002). Pode-se adotar duas abordagens para construção do *Data Warehouse*, cada uma destas é defendida por um autor:

- **Top-down:** estratégia defendida por Inmon, onde o DW é derivado dos *Data Marts*, indo do alto nível ao mais detalhado. É uma abordagem complexa e de alto risco de fracasso, pois é necessário uma visão completa da solução, onde primeiramente é construído todo o DW para uma posterior disponibilização dos DMs (ELIAS, 2014).

- **Bottom-up:** abordagem defendida por Kimball(2002), na qual os *Data Marts* são derivados do DW, ao passo que vai do menor nível até o maior. Ou seja, visa “dividir para conquistar”, sendo considerada uma estratégia conservadora. A construção de *Data Marts* é incremental até se chegar ao *Data Warehouse* completo, diminuindo assim a chance de fracasso do projeto. Outra grande vantagem desta abordagem é o retorno mais rápido, já que ocorrem entregas parciais (ELIAS, 2014).

Figura 4: Data Warehouse e abordagens top-down e bottom-up



Fonte: ELIAS (2014)

## 2.2 MODELO DIMENSIONAL

Segundo Kimball (1998), a mais importante diferença entre o modelo de dados de *Data Warehouse* e sistemas OLTP<sup>5</sup> está no modelo de dados. O grande ganho no sistema OLTP está nas execuções das transações, o que se deve a uma técnica chamada modelo Entidade Relacionamento (ER), que busca remover qualquer redundância de dados.

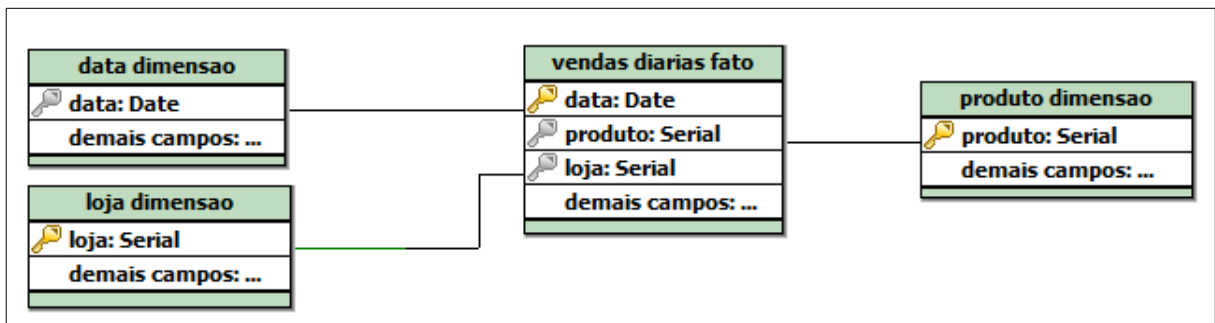
Ainda segundo Kimball (1998), o modelo Entidade Relacionamento é muito simétrico, todas as tabelas parecem iguais, não havendo meios de saber as maiores ou mais importantes, também não sendo possível identificar tabelas que possuem valores de negócios ou que armazenam descrições estáticas ou praticamente estáticas de objetos.

Ainda sobre o modelo ER, outro problema destacado pelo autor é que existe um número imenso de conexões possíveis entre duas tabelas, ou seja, o mesmo resultado é atingido através de diferentes meios. Em um banco de dados relacional isto é um sério problema, mesmo que diferentes consultas retornarem o mesmo resultado, o caminho escolhido faz toda a diferença.

O modelo dimensional defronta com o modelo Entidade Relacionamento, o qual é mais comumente utilizado entre os profissionais da TI. Pois este modelo começa com as tabelas, as quais pode ser (GOMES, 2010):

- **Tabela fatos:** Armazenam medições numéricas do negócio. Cada uma das medições é obtida na intersecção de todas as dimensões (KIMBALL,2012)
- **Tabela dimensão:** Armazenam as descrições textuais do negócio, onde cada descrição textual auxilia na definição de um componente. Em um banco de dados bem estruturado, estas tabelas possuirão muitos atributos (campos) (KIMBALL,2012).

Figura 5: Tabelas fato e dimensões



Fonte: Adaptado de KIMBALL (2002)

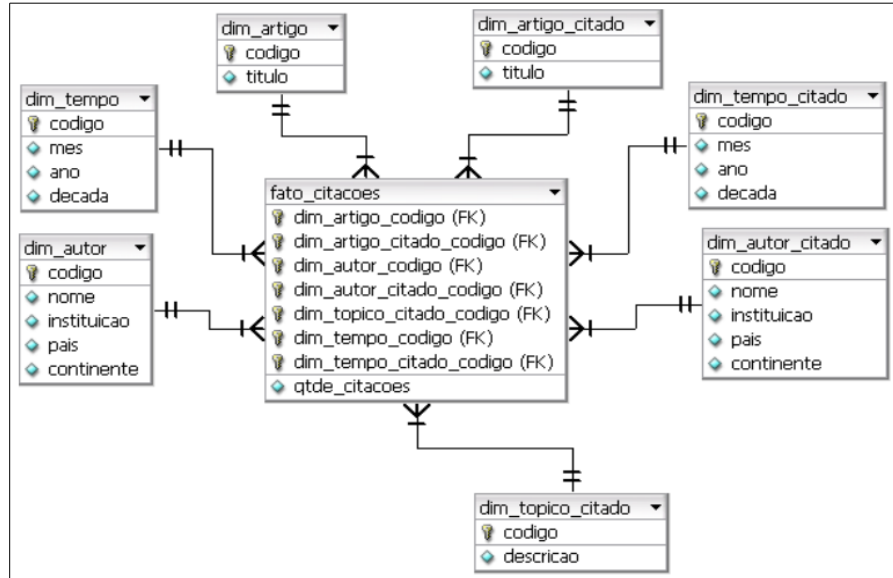
### 2.2.1 Esquema estrela

O esquema estrela é o modelo dimensional no qual existe uma tabela central (tabela

5 OLTP (*On-Line Transaction Processing*): Sistemas os quais registram todas as transações em tempo real, seja de inserção, remoção ou atualização (COSTA, 2012).

de fatos) contendo medidas quantitativas de natureza unitária ou transacional, e relacionada com múltiplas tabelas dimensões (OLIVA, 2015).

Figura 6: Esquema de estrela

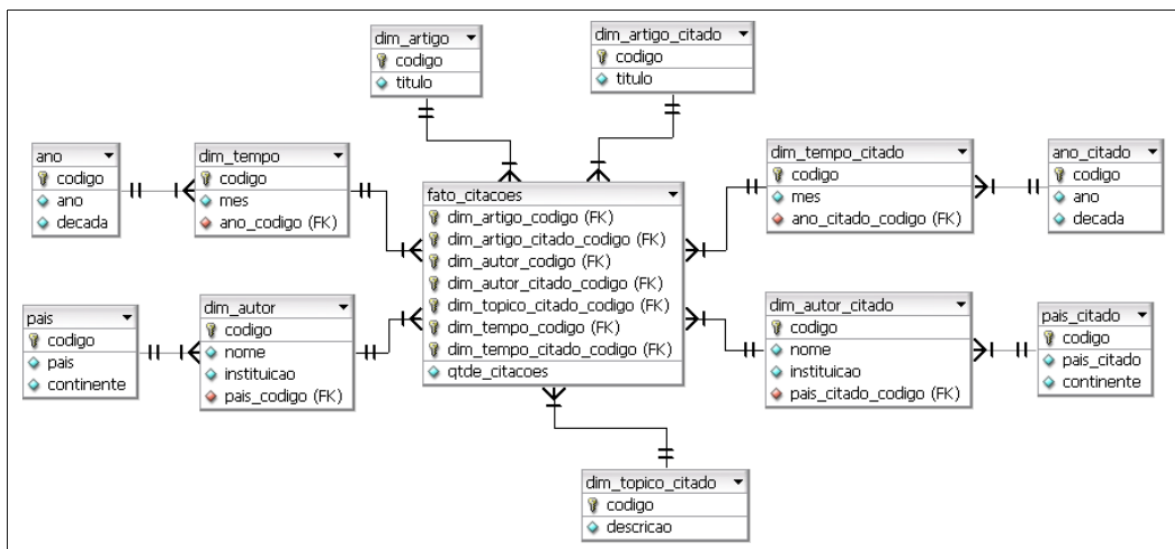


Fonte: KANASHIRO (2007)

### 2.2.2 Floco de neve

O esquema floco de neve é uma variação do esquema estrela, no qual as tabelas dimensões são organizadas em uma hierarquia. Deste modo a tabela de fatos não se relaciona diretamente com todas as dimensões (ELMASRI e NAVATHE, 2010).

Figura 7: Esquema floco de neve



Fonte: KANASHIRO (2007)

## 2.3 DATA MINING

Conforme Elmasri e Ramez (2010), como o próprio nome já diz, *Data Mining* refere-

se a garimpagem/mineração ou descoberta de novas informações com base em padrões ou regras em uma quantidade massiva de dados. *Data Mining* pode ser enquadrada como uma parte do processo de Descoberta de Conhecimento de Banco de dados (*Knowledge Discovery in Databases*), abreviada como KDD, o qual é composto por seis fases: seleção de dados, limpeza, enriquecimento, transformação ou codificação, *Data Mining* e construção de relatório e apresentação da informação.

De acordo com Rezende et al., (2003), o processo de *Data Mining* pode ser dividido em três etapas iterativas: pré-processamento dos dados, extração dos padrões e pós-processamento. Antes destas etapas, ocorre a identificação do problema, e após, a utilização do conhecimento captado (Figura 8).

Figura 8: Processo de *Data Mining*



Fonte: Rezende et al. (2003)

De acordo com Elmasri e Ramez (2010), para melhor utilização desta técnica, é essencial que os dados do *Data Warehouse* estejam agregados ou sumarizados, visto que esta técnica auxilia na obtenção de novas informações, as quais não poderiam ser identificadas através de *queries* (consultas) programadas. Os autores afirmam que o resultado da mineração pode ser o descobrimento de informações como:

**1. Regras de associação:** regras que relacionam a presença de um conjunto de itens com outro conjunto de itens variáveis. Por exemplo, uma pessoa que compra uma bola de futebol está propensa a comprar chuteiras.

**2. Hierarquia de classificação:** grupos ou classes de elementos, baseado em alguns parâmetros pré-definidos. Exemplos: classes sociais A, B e C.

**3. Padrões sequenciais:** seqüências de ações ou eventos associados a algum relacionamento temporal. Exemplo: Uma pessoa que esteve endividada está propensa a estar endividada novamente dentro de algum período de tempo.

**4. Padrões com séries temporais:** Similaridades encontradas em períodos de tempos. Por exemplo, produtos que vendem mais durante o verão.

**5. Clustering (Agrupando):** elementos que podem ser segmentados em conjuntos de elementos “similares”. Um exemplo é a população brasileira que pode ser categorizada em diversos grupos de propensão de compra, desde os mais propensos até os menos propensos.

Segundo Primak (2008), *Data Mining* além de ser um processo complexo, exige não somente um especialista técnico em mineração de dados, mas também profissionais com conhecimento no negócio. Isso para evitar que toda a modelagem dos dados tenha sido feita em vão e assim mais facilmente se alcançar os resultados e descobertas de informações citadas.

## 2.4 APLICAÇÕES FRONT-END

Com o crescimento de tecnologias e dos conceitos de BI surgiu uma enorme gama de aplicações *front-end*. Tais ferramentas conseguem ser melhor aceitas, por terem uma interface amigável e boa usabilidade perante diversos níveis e nichos de usuários, conseguindo assim atingir profissionais de vários setores, não apenas diretores e gerentes ou profissionais de TI. Estas ferramentas inclusive conseguem propôr aos usuários diferentes visões de uma informação sem a necessidade de requisitar algum profissional capacitado para isto (PRIMAK, 2008).

Conforme Costa (2012), um fator determinante para sucesso do BI é a elaboração de uma boa camada de apresentação, onde em um único painel a informação relevante esteja disponível. Dependendo do tipo de análise que cada usuário pretende fazer, a ferramenta de visualização mais indicada pode envolver (SILVA, 2011):

- **Cubos OLAP:** Gerados a partir dos dados dimensionais disponíveis no DW e podem ser explorados por meio de tabelas *pivot*<sup>6</sup>. Permitem que o usuário execute operações

6 *Pivot:* Mover um objeto das linhas para as colunas e vice-versa ou alterar a ordem dos objetos nas colunas ou nas linhas num relatório para ver os dados de diferentes perspectivas. (GOUVEIA, 2013)

de consulta como *slicing* e *dicing* (fatiamento de dados), *drill down/up* (navegação detalhada/sumarizada através das dimensões do cubo) e *roll-up* (agregações de diferentes níveis do cubo) (SILVA, 2011).

- **Scorecards:** É um tipo de relatório que mede o desempenho empresarial em relação a determinados objetivos usando indicadores KPIs (*Key Performance Indicator* – Indicadores-chave de desempenho). Cada KPI representa um aspecto do desempenho empresarial e está sempre associado a um objetivo. A principal função do *scorecard* é mostrar se a empresa está cumprindo ou falhando no seu planejamento estratégico (GOUVEIA, 2013)

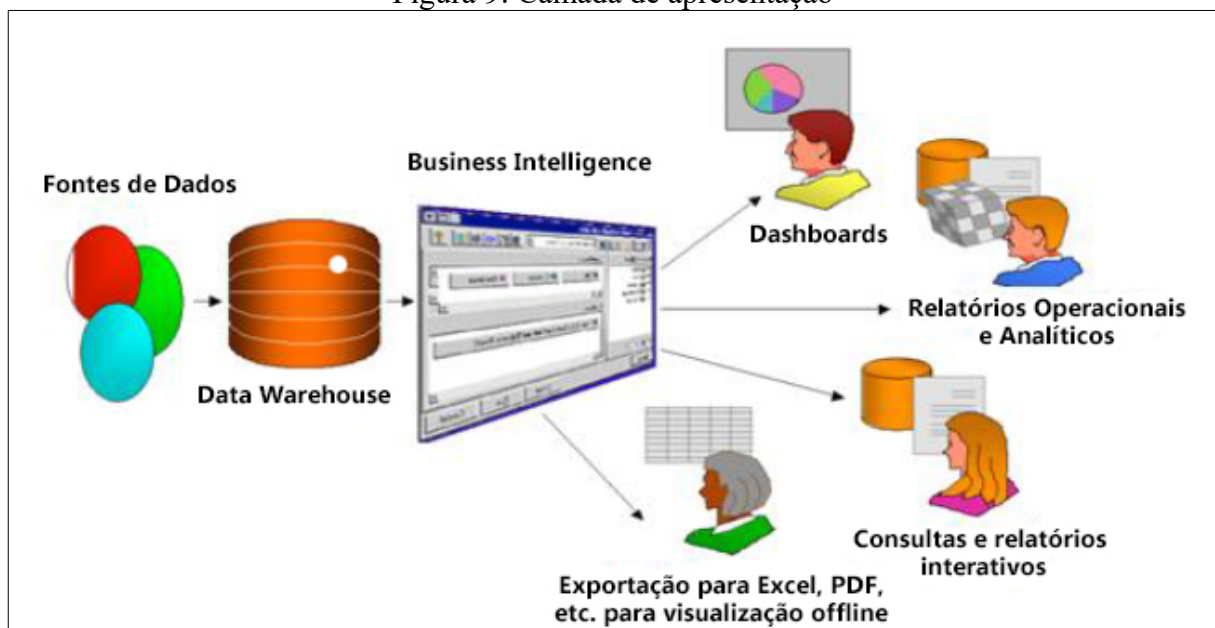
- **Dashboards:** Representação de diversos KPIs em um único painel sumarizado, facilitando a visão macro dos indicadores (SILVA, 2011). É um painel de informação que oferece aos utilizadores um alto nível de interatividade, permitindo ao usuário combinar vários componentes (tabelas, gráficos, *scorecards*, entre outros) de forma a explorar os dados com capacidades OLAP como *drill* e o *slice and dice* (GOUVEIA, 2013).

- **Relatórios:** Representado através de gráficos e/ou tabelas, dependendo do foco (operacional ou analítico), para apresentar maiores detalhes a cerca de um contexto (SILVA, 2011).

- **Alertas:** Disparadores de avisos para os responsáveis sobre certas condições pré-definidas (SILVA, 2011).

A Figura 9 apresenta a camada de apresentação de um sistema para BI.

Figura 9: Camada de apresentação



Fonte: Adaptado de SILVA (2011)



## 2.5 TRABALHOS RELACIONADOS

Para a construção deste estudo, foram analisados uma série de artigos a cerca de *Data Warehouse*, *Business Intelligence* e seus resultados.

Quadro 3: Trabalhos relacionados

(continua)

<b>Autor</b>	<b>Título</b>	<b>Objetivo</b>
Silva (2012)	Uma abordagem utilizando <i>Business Intelligence</i> para apoiar o processo de tomada de decisão na gestão da evolução de serviços web.	Propor uma abordagem de BI para apoio à tomada de decisão de um provedor de serviços durante a evolução de serviços em um portfólio.
Martiny e Ribeiro (2015)	Módulo para <i>Business Intelligence</i> Compatível com Dispositivos Móveis para o Software de Gestão Empresarial Elementare;	Construir um módulo para <i>Business Intelligence</i> compatível com dispositivos móveis para o ERP Elementare <sup>7</sup> .
Coelho et. al. (2012)	<i>Business intelligence</i> : suporte à gestão do desempenho científico.	Formalizar uma estrutura de BI, composta de recursos como ETL, DW, DM, <i>Data Mining</i> e OLAP para dar suporte a gestão do desempenho científico.
Cayres, Oliveira e Marini (2009)	<i>Business Intelligence</i> na era da informação e as vantagens do Oracle na efetivação dessa tecnologia.	Apresentar e demonstrar principais características e necessidades da tecnologia para sistema de BI. Sedimentar os elementos através do Oracle <sup>8</sup> para a efetivação <i>Business Intelligence</i> .
Paula (2016)	<i>Business intelligence</i> no auxílio da gestão da inovação: um estudo de caso utilizando <i>sql server integration services</i> e <i>microstrategy</i> .	Descrever características de duas ferramentas que auxiliam no desenvolvimento de projeto de BI: <i>SQL Server Integration services</i> e <i>Microstrategy</i> <sup>9</sup> .
Almeida e Carmago (2015)	Aplicando Técnicas de <i>Business Intelligence</i> sobre dados de desempenho Acadêmico: Um estudo de caso.	Aborda uma implementação de <i>Business Intelligence</i> dentro de uma instituição federal de ensino superior, a fim de otimizar a aplicação de seus recursos.

7 ERP Elementare: <http://www.elementare.inf.br/> acessado em 30/11/2017.

8 Oracle: <https://www.oracle.com/br/index.html> acessado em 30/11/2017.

9 Microstrategy: <https://www.microstrategy.com/br> acessado em 30/11/2017.



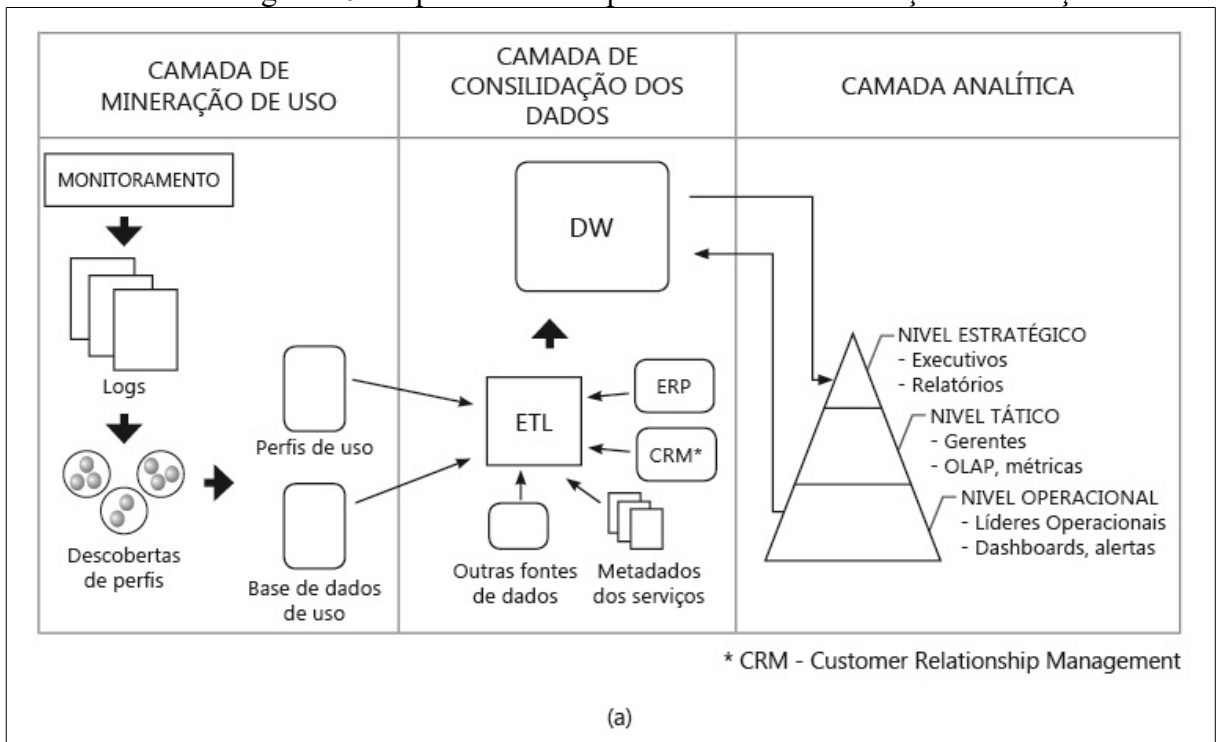
(conclusão)

Autor	Título	Objetivo
Fortulan e Filho (2005)	Uma proposta de aplicação de <i>Business Intelligence</i> no chão-de-fábrica.	Propor a implantação de um sistema para BI por meio de ferramentas de <i>Data Warehouse</i> e OLAP, aplicadas ao chão-de-fábrica. Desenvolver um sistema que utilize os dados resultantes do processo produtivo e os transforme em informações que auxiliem na tomada de decisões.

Fonte: O Autor (2018)

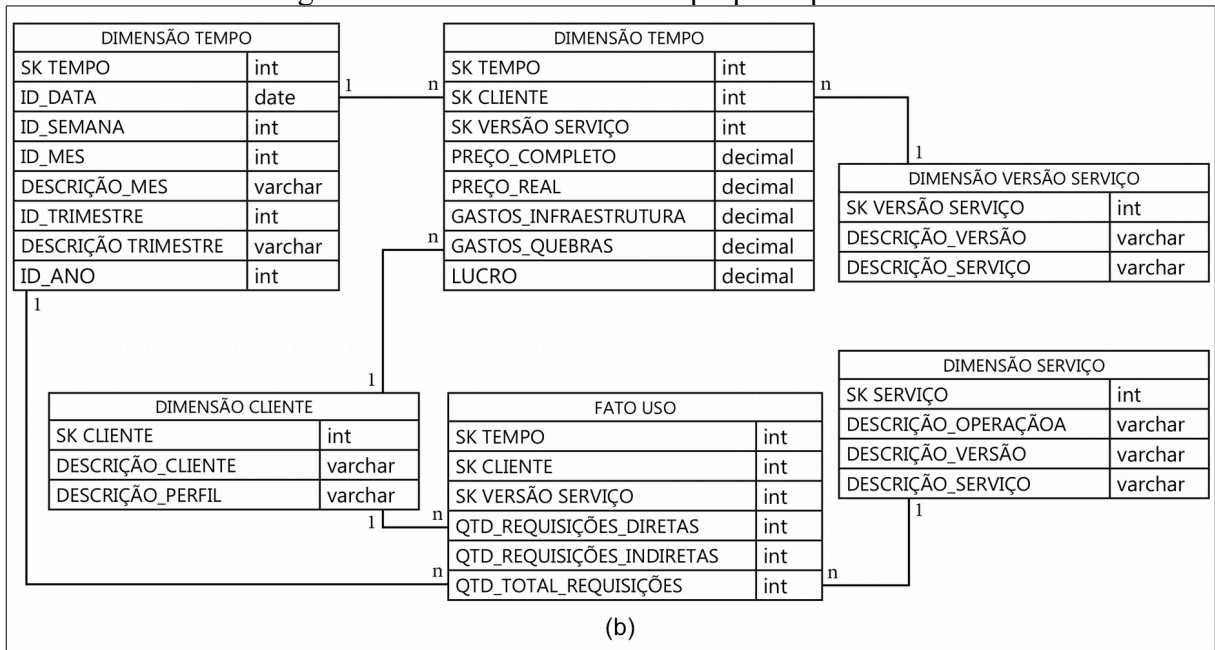
O trabalho de Silva (2012) alcançou o resultado em conjunto com outro trabalho, focado na mineração do uso de serviços para geração das bases de perfis e de dados de uso, resultados os quais estão relatados em Silva et al. (2012). Silva (2012) identificou ser possível inserir o BI no apoio à tomada de decisões do provedor de serviços. Apresentou em sua obra uma arquitetura de BI, composta por Camada de Mineração do Uso, Camada de Consolidação e Camada Analítica, as quais são ilustradas junto ao modelo dimensional construído e apresentado na Figura 10 e 11.

Figura 10: Arquitetura de BI para o cenário de evolução de serviços



Fonte: Adaptado de SILVA (2012)

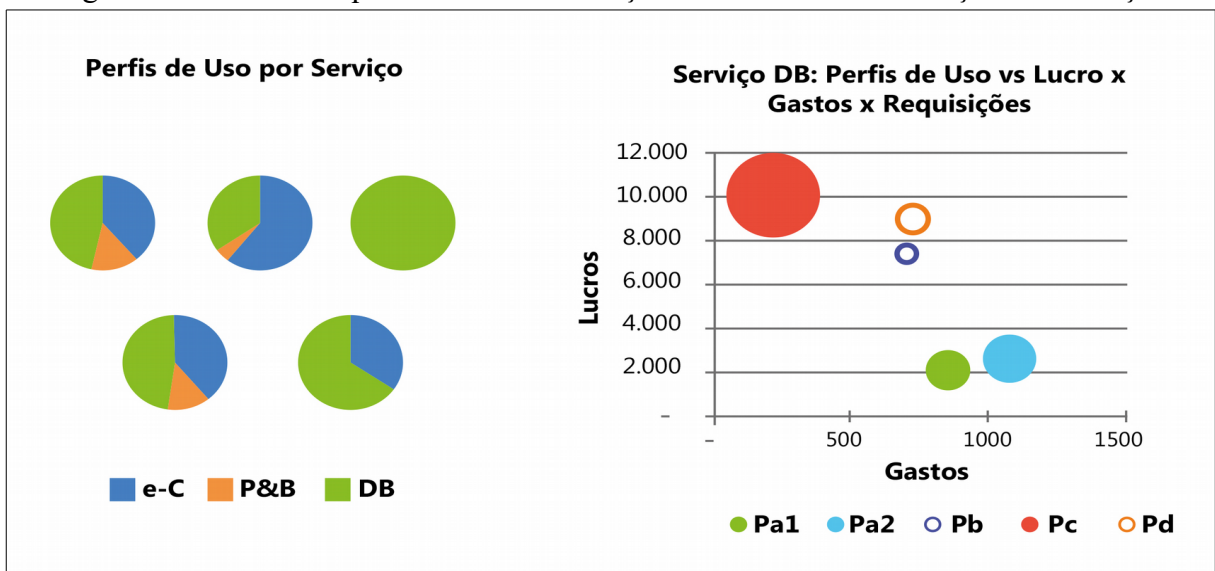
Figura 11: Modelo dimensional proposto para o DW



Fonte: Adaptado de SILVA (2012)

Ainda a respeito de seu trabalho, Silva (2012) realizou uma prova de conceito deste BI e construiu um *dashboard* (Figura 12) para evidenciar as métricas que foram propostas. Neste é possível visualizar gráficos de pizza, representando a proporção de requisições por serviço para cada grupo de usuários. Também é possível identificar um gráfico de bolha o qual relaciona informações sobre Lucro x Gastos x Requisições com cada perfil de uso e um serviço específico. Através da prova de conceito foi possível identificar mais pontos que devem ser analisados, definidos e detalhados em sua abordagem.

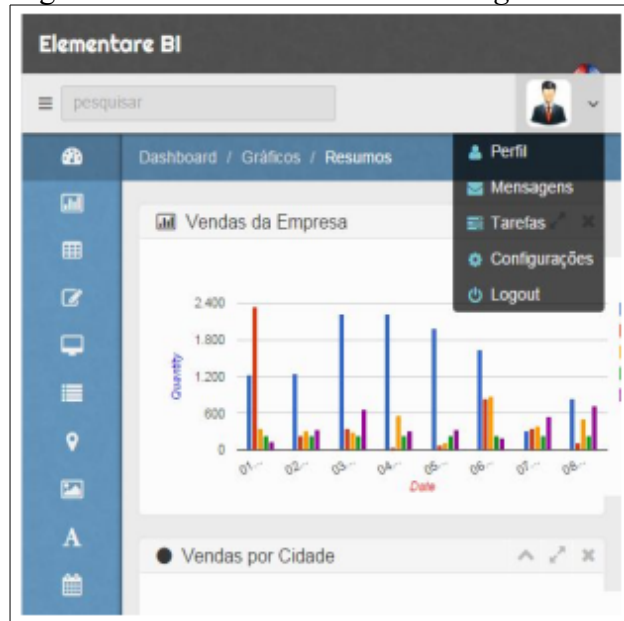
Figura 12: Dashboard que relaciona informações sobre custos e utilização dos serviços



Fonte: Adaptado de SILVA (2012)

No trabalho de Martiny e Ribeiro (2015), para atingir seus resultados foi necessário a construção de um *Data Warehouse* e realizar o ETL para popular o banco de dados (MySQL<sup>10</sup>) da aplicação *Web*. A camada *front-end* foi desenvolvida com *framework Bootstrap*<sup>11</sup> com bibliotecas de gráficos *Morris Charts*<sup>12</sup> e *Google Charts*<sup>13</sup>. A Figura 13 destaca o Layout do *Elementare Business Intelligence*, construído para dispositivos móveis, em formato responsivo, o qual permite o acesso de diferentes plataformas móveis e em diversos tamanhos de tela. Utilizando bibliotecas que permitem a geração de vários tipos de gráficos configuráveis de acordo com os parâmetros enviados, e que montam automaticamente os gráficos.

Figura 13: *Dashboard* utilizando *Google Charts*



Fonte: MARTINY e RIBEIRO (2016)

De acordo com o artigo, o software não está finalizado e está em fase de testes com algumas empresas. Algumas funcionalidades, tais como gráficos *drill-down*, detalhes do envio de relatórios por e-mail não estavam concluídas.

Coelho et. al. (2012), apresentou em seu artigo um modelo de BI para auxílio aos gestores do desempenho científico do Instituto Federal do Espírito Santo (IFES), sugerindo a utilização de um sistema ETL para alimentar o DW e os *Data Marts* e as análises destes dados encarregados de serem executadas através de *Data Mining* e OLAP. Nas conclusões dessas análises, sugeriu-se emitir relatórios com os resultados em forma de planilha, gráfico e tabelas.

10 MySQL: <https://www.mysql.com/> acessado em 30/11/2017.

11 Bootstrap: <https://getbootstrap.com/> acessado em 30/11/2017.

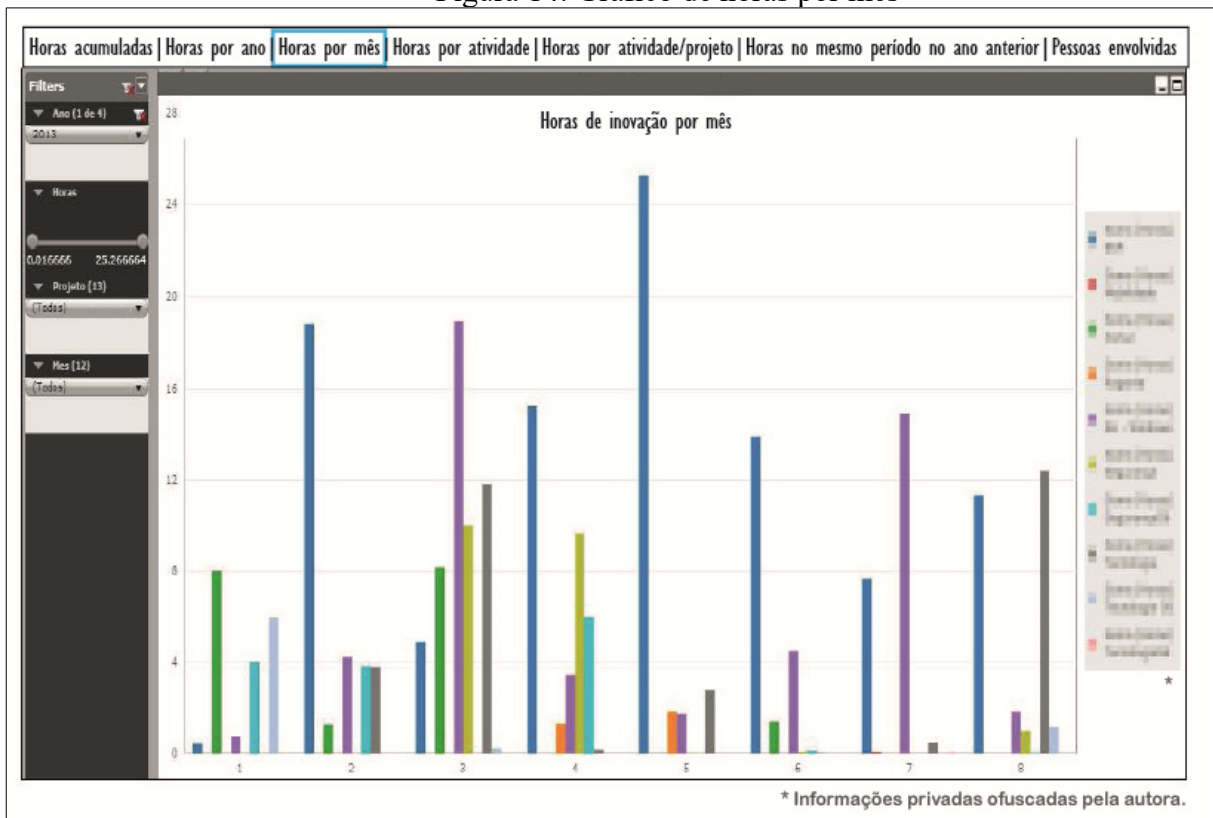
12 Morris Charts: <http://morrisjs.github.io/morris.js/> acessado em 30/11/2017.

13 Google charts: <https://developers.google.com/chart/> acessado em 30/11/2017.

Cayres, Oliveira e Marini (2009), abordaram o quanto o BI é fundamental para o sucesso das organizações. Sustentaram como a Oracle se apresenta como um dos principais *players* no mercado, possuindo em seu portfólio todas as ferramentas necessárias que vão desde ferramentas de buscas pelos dados dos sistemas transacionais, ferramentas de ETL, a alocação do seu conteúdo nos bancos de dados e a disponibilização das informações de forma gráfica. Afirmam ainda as vantagens da ferramenta *Oracle Warehouse Builder* e sua integração entre os demais componentes do banco de dados.

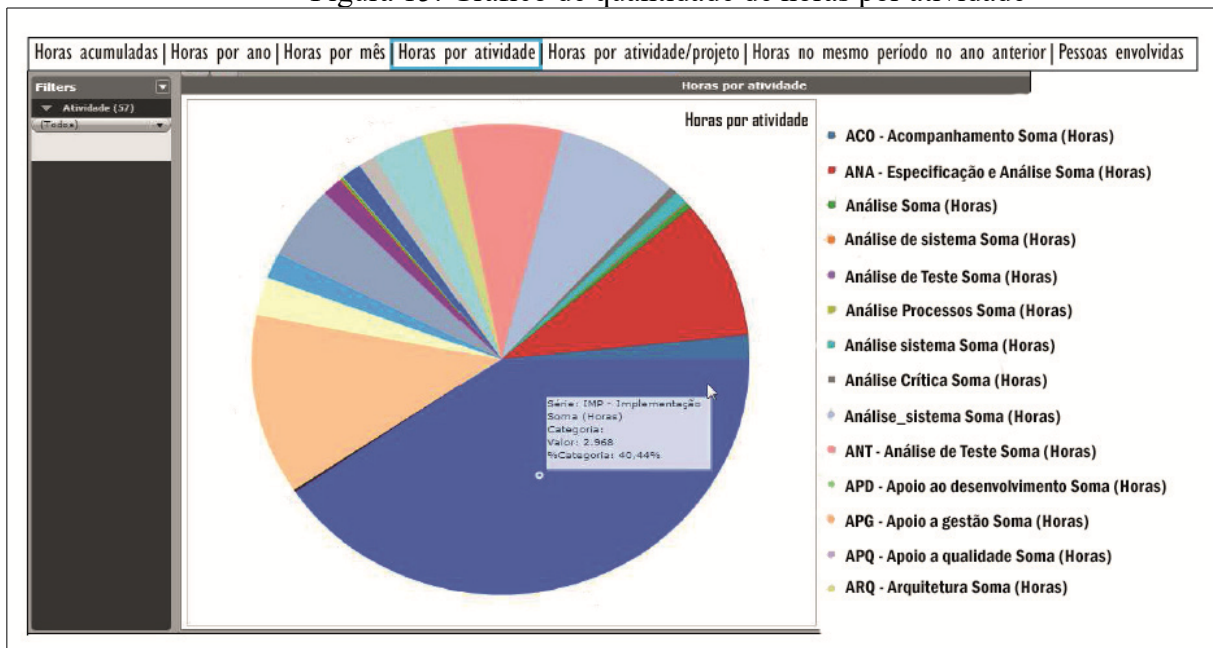
Paula (2016), apresenta uma solução para apoiar a gestão da empresa, onde também foi possível automatizar alguns processos. Outro item implementado neste trabalho foi um *dashboard*, centralizando várias informações que anteriormente estavam espalhadas por diversos documentos e arquivos. Também apresentou características a cerca da plataforma de integração de dados: *SQL Server Integration Service* (SSIS), e da ferramenta MicroStrategy. As Figuras 14 e 15 representam os gráficos construídos.

Figura 14: Gráfico de horas por mês



Fonte: PAULA (2016)

Figura 15: Gráfico de quantidade de horas por atividade



Fonte: PAULA (2016)

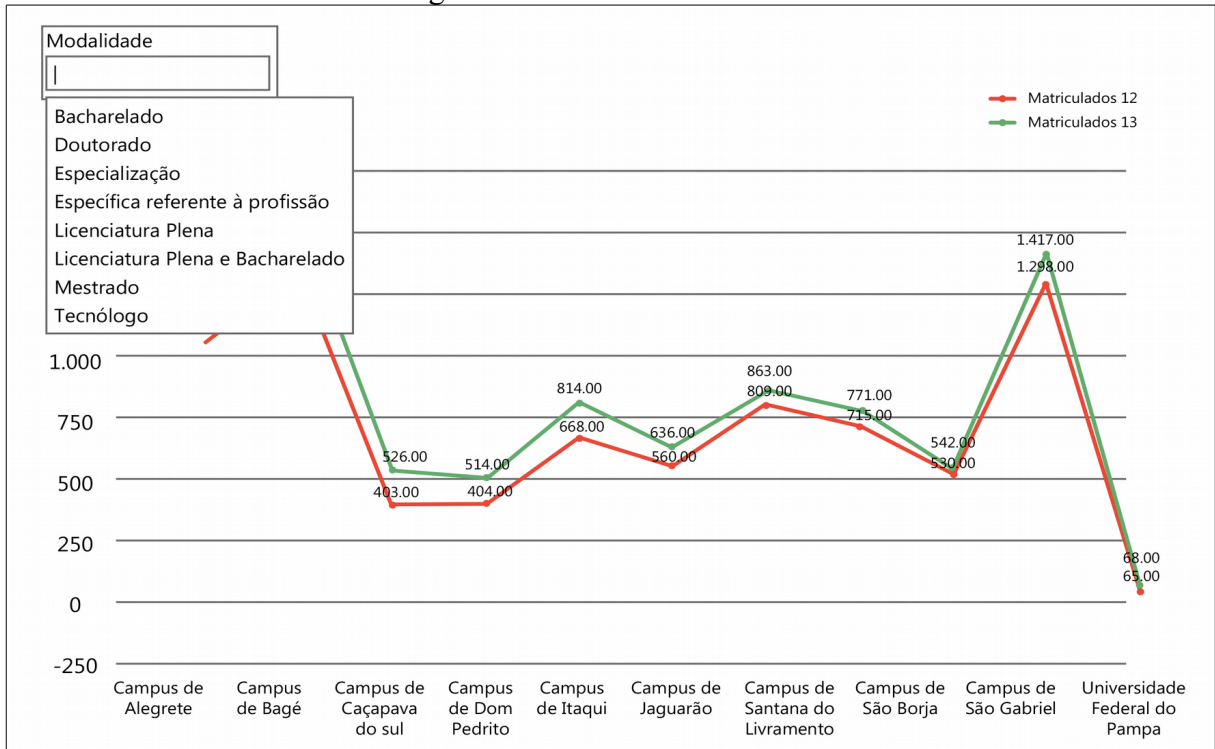
A autora desenvolveu sua solução visual através da ferramenta MicroStrategy Web. As consultas implementadas possuem filtros de busca e opção de exportação para PDF ou Excel. Paula (2016), apresentou em seu artigo uma consulta de horas por mês, em forma de gráfico de barras (Figura 14), onde é possível ver a soma de horas por projeto ou por mês e possui filtros por ano, horas, projeto e mês. Também apresentou um gráfico de pizza, para representar a quantidade de horas por atividade (Figura 15).

Almeida e Camargo (2015), decidiram por implementar um DM acadêmico, o qual é possível realizar consultas como: Avaliações sobre os cursos de graduação e pós-graduação; Análise dos cursos com maiores índices de evasão; Análise sobre número de formandos em todos os campus; Análises com o número de matrículas por campus, podendo verificar aumentos ou diminuições em relação aos últimos anos.

Com a implantação do *Data Mart* os autores utilizaram a plataforma SpagoBI<sup>14</sup> para montar painéis gerenciais e desta forma construíram um gráfico de linhas, que representam as matrículas referentes a cada campus da UNIPAMPA, permitindo ao usuário informar o filtro de modalidade. Também apresentaram um *dashboard* construído com base nos atributos números de alunos inscritos, números de alunos formados por campus e números de alunos evadidos a cada curso da instituição. Painéis que foram base de suporte à decisão sob uma das áreas de interesse da instituição (Figura 16 e 17):

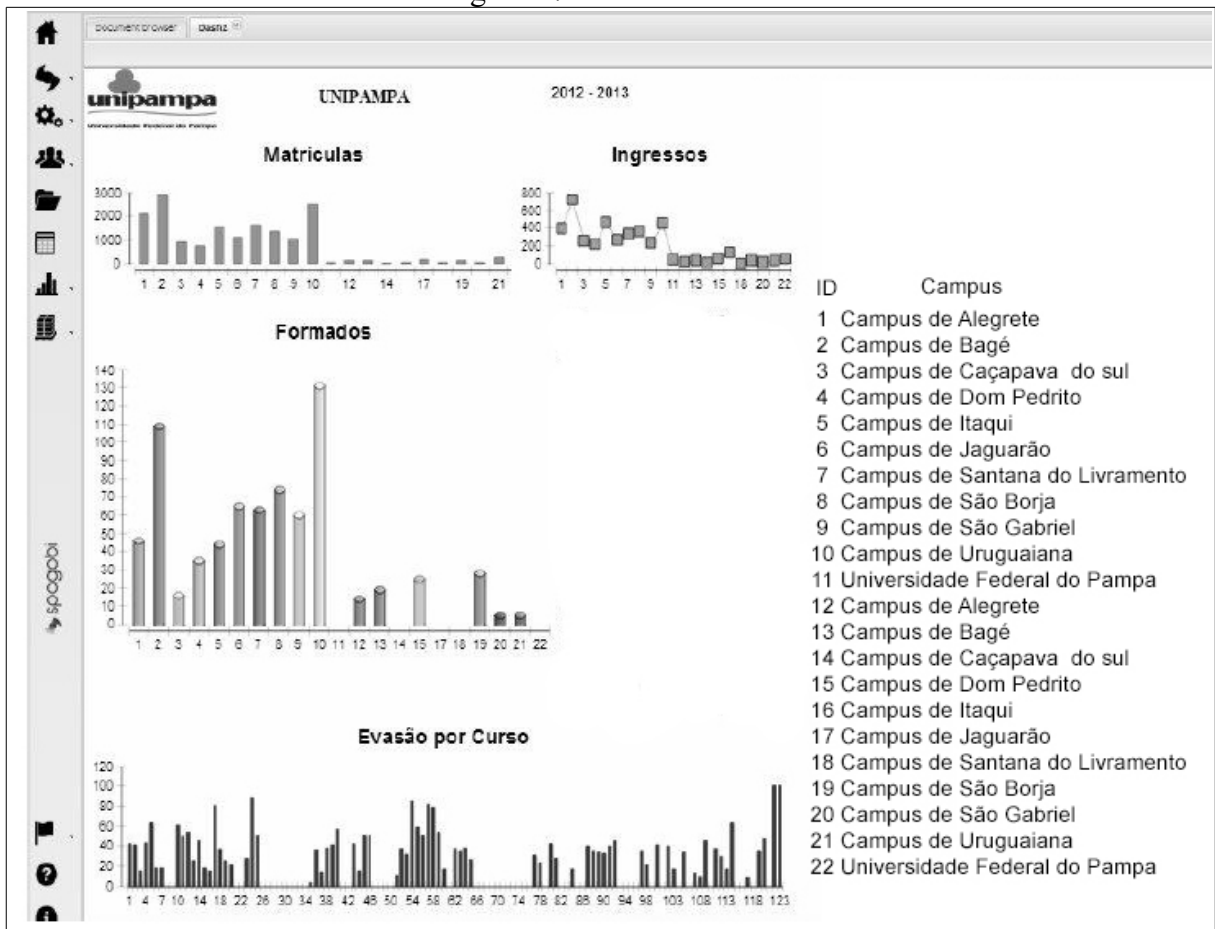
14 SpagoBI: <https://www.spagobi.org/> acessado em 30/11/2017.

Figura 16: Gráfico de matrículas



Fonte: Adaptado de ALMEIDA e CAMARGO (2015)

Figura 17: Dashboard UNIPAMPA

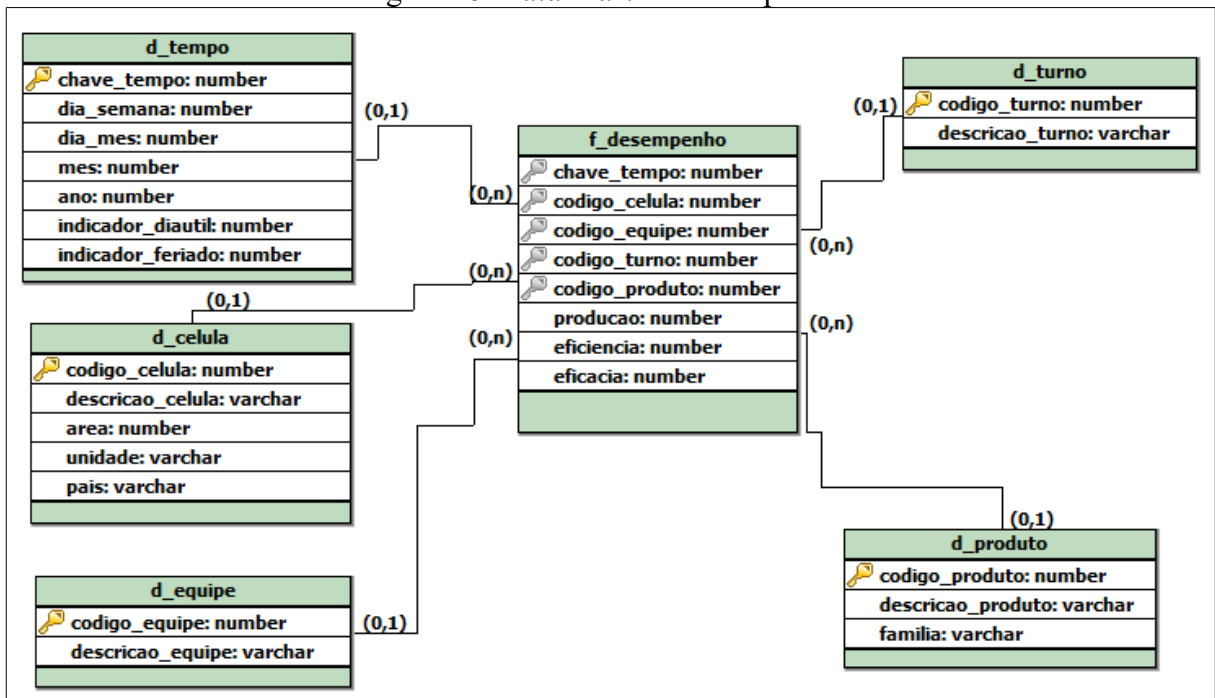


Fonte: Adaptado de ALMEIDA e CAMARGO (2015)



Fortulan e Filho (2005), relataram o quanto os dados gerados no chão-de-fábrica normalmente ficam espalhados por diversas fontes, como diferentes bancos de dados e folhas de papel, dificultando assim o levantamento e recuperação de informações. Apresentaram também os benefícios do uso do BI e *Data Warehouse*, elencando vantagens do uso destas ferramentas. Para realizar os testes do modelo proposto pelos autores, foi utilizado o banco de dados *Microsoft Access*<sup>15</sup> e *Analysis Services*<sup>16</sup>. A Figura 18 apresenta o *Data Mart* de Desempenho elaborado neste projeto, onde pode ser identificado o esquema estrela.

Figura 18: *Data Mart* de desempenho modelado



Fonte: Adaptado de Fortulan e Filho (2005)

## 2.6 CONSIDERAÇÕES FINAIS

Neste capítulo foi apresentado conceitos de BI, desde sua evolução, sua arquitetura e sua importância. Também sobre *Data Warehouse*, abordando sua definição, suas subdivisões em *Data Marts*, seu modelo dimensional e seus esquemas, elencando também conteúdo a respeito de *Data Mining* e os resultados obtidos com essa técnica, e sobre aplicações *front-end*, exemplificando através de algumas soluções conhecidas. Também foram apresentados uma série de trabalhos relacionados, dentre alguns pode-se extrair uma série de características a cerca de formas de visualização dos dados, apresentadas no Quadro 4:

15 *Microsoft Access*: <https://products.office.com/pt-br/access> acessado em 30/11/2017.

16 *Analysis Services*: <https://docs.microsoft.com/pt-br/sql/analysis-services/analysis-services> acessado em 30/11/2017.

Quadro 4: Características visuais

<b>Autor</b>	<b>Gráfico</b>			
	<b>Linhas</b>	<b>Pizza</b>	<b>Bolha</b>	<b>Barras</b>
Silva (2012)		X	X	
Martiny e Ribeiro (2016)				X
Paula (2016)		X		X
Almeida e Camargo (2015)	X			X

Fonte: O Autor (2018)

Os artefatos que foram explanados no Quadro 4, auxiliarão a propor uma forma de visualização dos dados, bem como os demais itens que foram citados e serão tomados como base para fundamentar a proposta de solução, que será descrita no próximo capítulo.



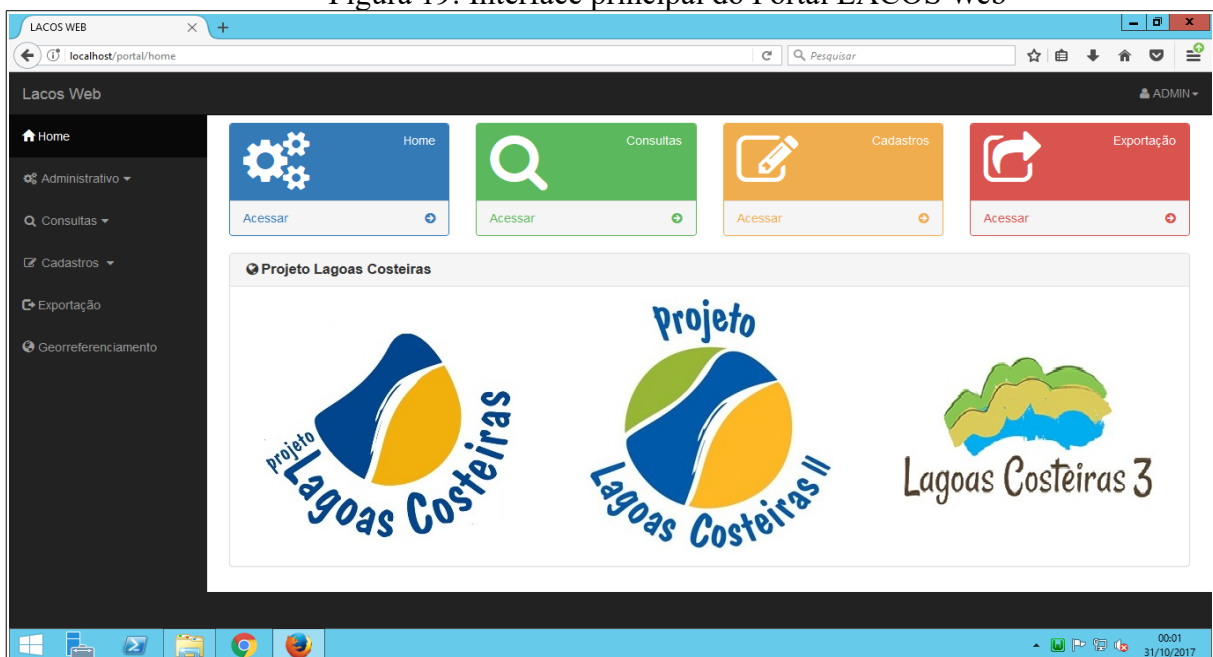
### 3 PROPOSTA DE SOLUÇÃO

A proposta de solução para o presente projeto será definido sobre recursos e algumas técnicas que já foram utilizadas pelo portal LACOS Web, seguindo os padrões tecnológicos, para assim facilitar a aceitação e adequação dos usuários. Para isto serão utilizados os mesmos artefatos já utilizados pelo restante do portal. Além do levantamento de requisitos junto aos pesquisadores para tomar conhecimento a cerca de quais métricas são importantes serem exibidas em um *dashboard* de visualização. Portanto, este capítulo é embasado sobre a construção de um *Data Warehouse*, propondo como será o ETL, até as formas como serão visualizadas as consultas sobre esse banco de dados.

#### 3.1 SITUAÇÃO ATUAL

Atualmente no portal LACOS Web, as funcionalidades para georreferenciamento das lagoas, cadastros básicos, consultas, importação e exportação de dados já foram entregues e estão sendo usadas pelos pesquisadores. Sendo assim, a modelagem do banco de dados também já foi realizada. A Figura 19 apresenta a página principal do Portal LACOS Web.

Figura 19: Interface principal do Portal LACOS Web



Fonte: O Autor (2018)

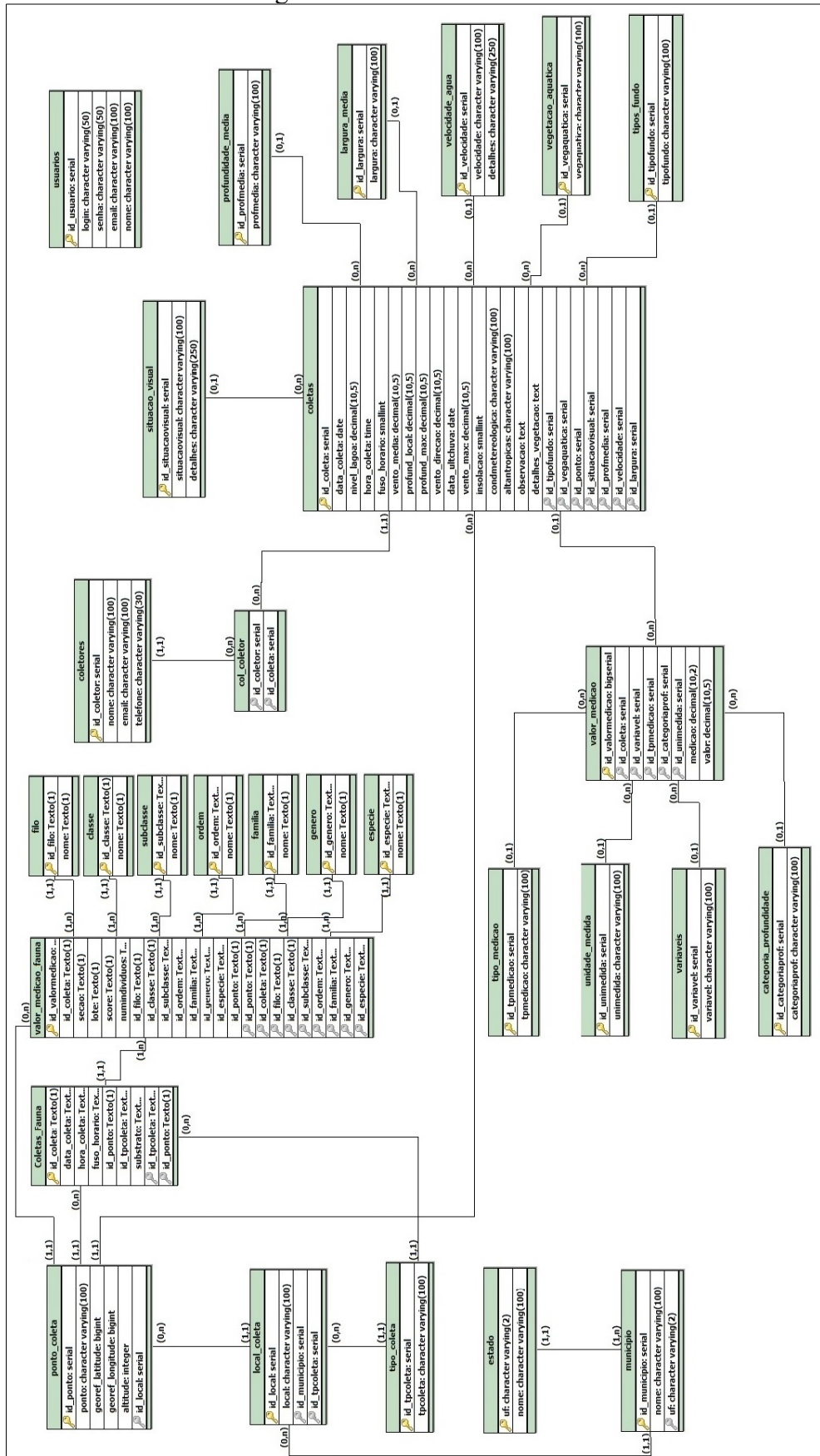
O portal foi desenvolvido na linguagem de programação PHP<sup>17</sup>, rodando em servidor de aplicação Wamp<sup>18</sup>. O banco de dados foi construído no PostgreSQL<sup>19</sup> e seu modelo lógico pode ser observado pela Figura 20:

17 PHP: <http://php.net/> acessado em 30/11/2017.

18 Wamp: <http://www.wampserver.com/en/> acessado em 30/11/2017.

19 PostgreSQL: <https://www.postgresql.org/?&> acessado em 30/11/2017.

Figura 20: Modelo ER do LACOS Web



Fonte: O Autor (2018)

### 3.2 REQUISITO NECESSÁRIOS

A engenharia de software é de suma importância pois adota uma abordagem sistemática e organizada, a qual, é a maneira mais eficiente de se produzir um software de qualidade. Para isso reúne premissas de engenharia, as quais aplicam teorias, métodos e ferramentas para fazer algo funcionar. Com aspectos de produção de software, que estão ligados a meios técnicos do desenvolvimento do produto e gerenciamento de projeto (SOMMERVILLE, 2011).

Requisitos devem ser a descrição de serviços fornecidos pelo sistema e suas restrições operacionais, os quais devem refletir as necessidades do cliente a resolver algum problema.

Segundo Filho (2009), pode-se dividir os requisitos em duas características especiais:

- **Funcionais:** Representam os comportamentos que um programa ou sistema deve apresentar diante de certas ações de seus usuários (FILHO, 2009).
- **Não funcionais:** Quantificam determinados aspectos do comportamento (FILHO, 2009).

Os próximos subcapítulos detalharão cada requisito para este projeto, os quais foram concebidos com base em padrões de construção de *Data Warehouse* e na leitura dos trabalhos relacionados.

#### 3.2.1 Requisitos funcionais do sistema

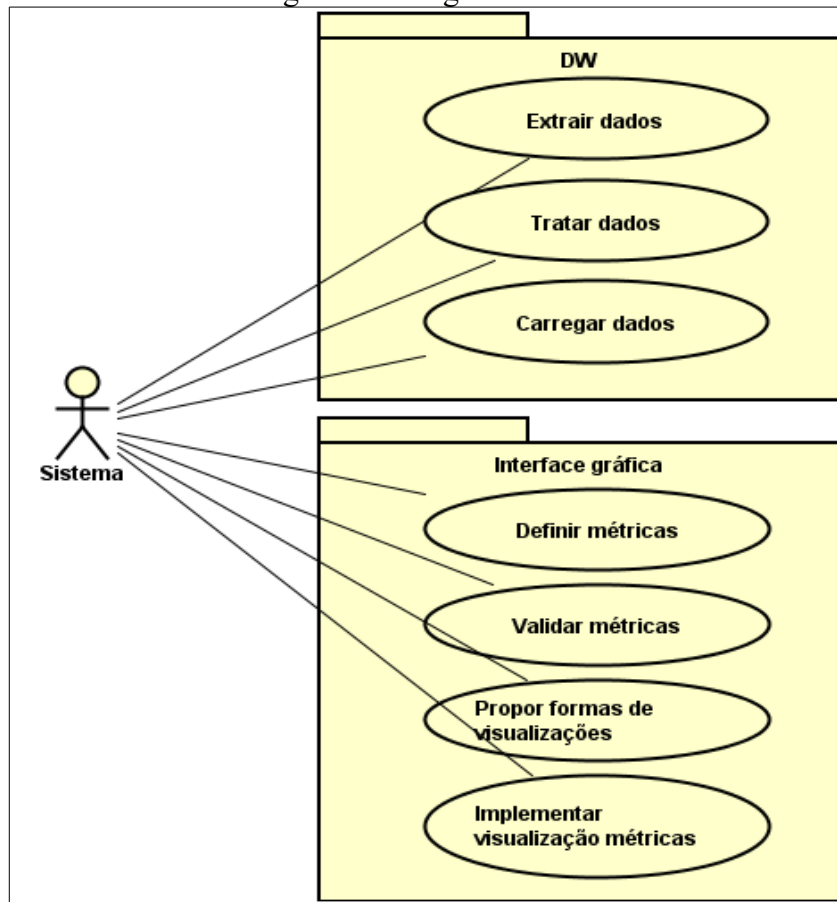
Sendo assim, os requisitos funcionais identificados para este projeto, são os dispostos no Quadro 5 e através do diagrama da Figura 21, os quais serão melhor elucidados nos próximos subcapítulos.

Quadro 5: Requisitos funcionais do sistema

Identificador	Descrição
<b>RF01</b>	Projetar a estrutura do <i>Data Warehouse</i>
<b>RF02</b>	Realizar a extração dos dados
<b>RF03</b>	Realizar o tratamento dos dados
<b>RF05</b>	Realizar a carga dos dados no <i>Data Warehouse</i>
<b>RF06</b>	Definir métricas
<b>RF07</b>	Validar métricas junto aos pesquisadores
<b>RF08</b>	Propor forma de visualização
<b>RF09</b>	Implementar forma de visualização de métricas

Fonte: O Autor (2018)

Figura 21: Diagrama de caso de uso



Fonte: O Autor (2018)

### 3.2.2 Requisitos funcionais do usuário

Os requisitos funcionais, em nível de usuário final da aplicação, estão dispostos no Quadro 6 e na Figura 22:

Quadro 6: Requisitos funcionais do usuário

(Continua)

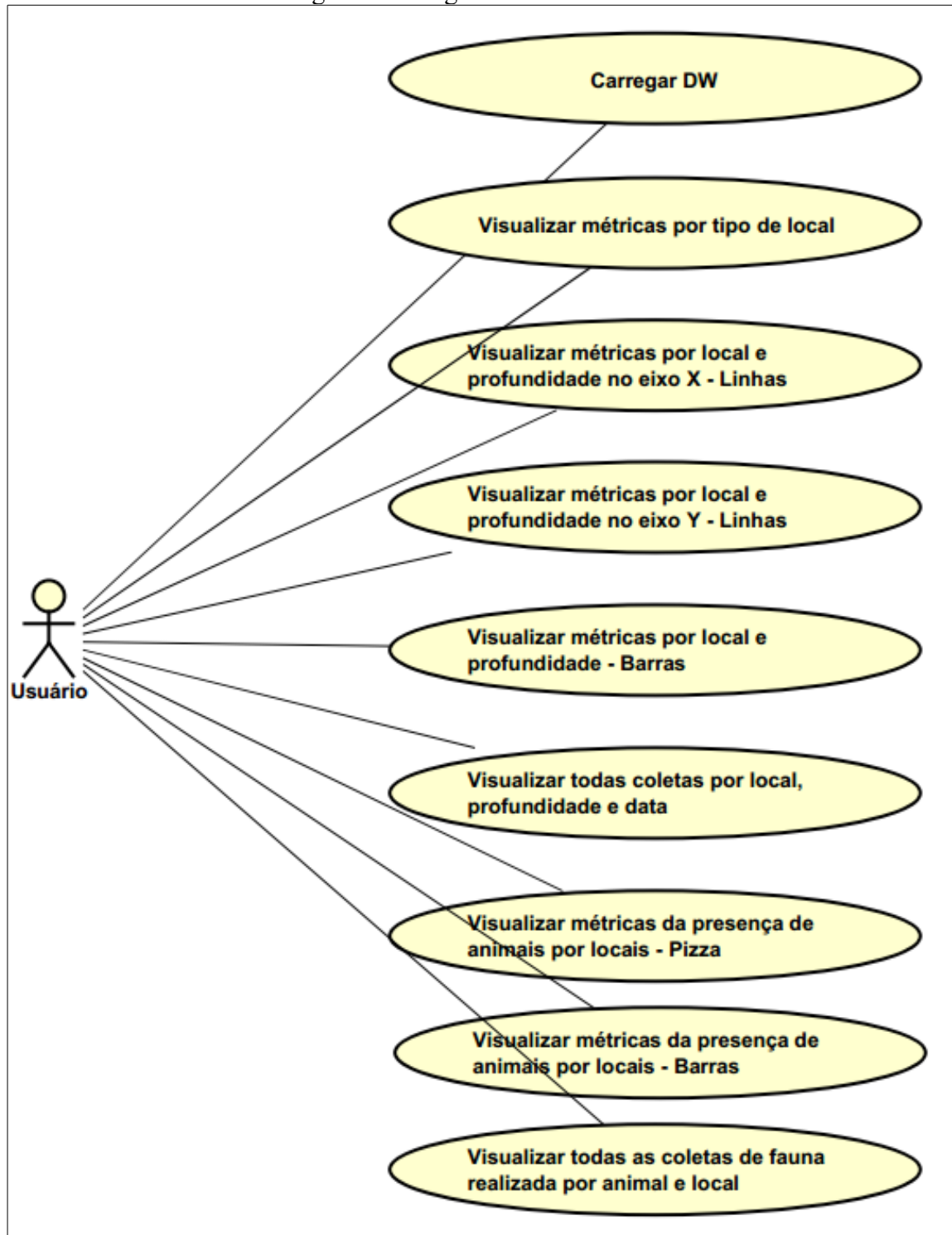
Identificador	Descrição
<b>RFU1</b>	Carregar DW
<b>RFU2</b>	Visualizar métricas físico/químico por tipo de local
<b>RFU3</b>	Visualizar métricas físico/químico por local com profundidade no eixo X – Gráfico em linhas
<b>RFU4</b>	Visualizar métricas físico/químico por local com profundidade no eixo Y – Gráfico em linhas
<b>RFU5</b>	Visualizar métricas físico/químico por local – Gráfico em barras
<b>RFU6</b>	Visualizar todas as coletas físico/químico realizadas por local, profundidade e data

(Conclusão)

Identificador	Descrição
<b>RFU7</b>	Visualizar métricas de presença de animais por local – Gráfico em pizza
<b>RFU8</b>	Visualizar métricas de presença de animais por local – Gráfico em barras
<b>RFU9</b>	Visualizar todas as coletas de fauna realizadas por animal e local

Fonte: O Autor (2018)

Figura 22: Diagrama do caso de uso



Fonte: O Autor (2018)

### 3.2.3 Requisitos não funcionais

Os requisitos não funcionais para o projeto são os destacados no Quadro 7:

Quadro 7: Requisitos não funcionais

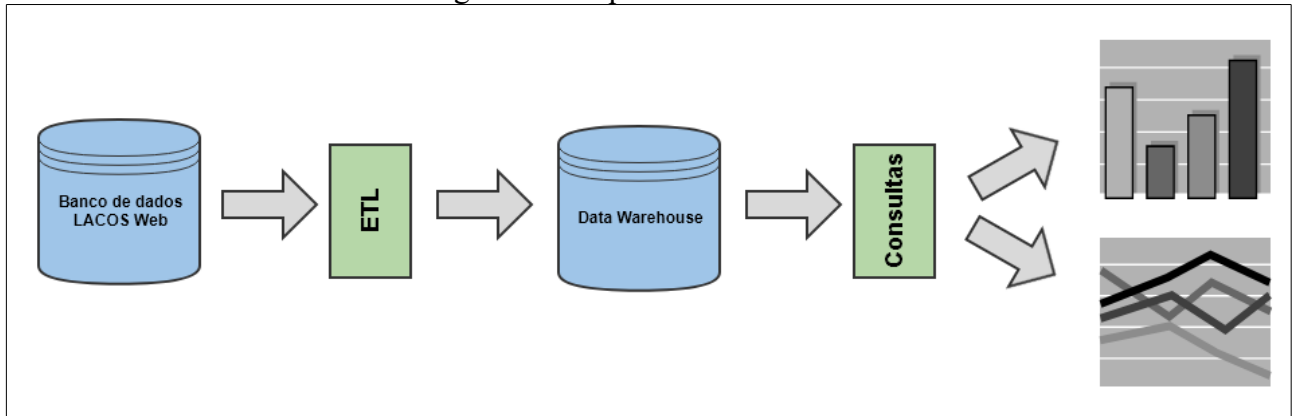
Identificador	Descrição
NF01	Sistema LACOS Web e banco de dados estarem funcionando corretamente
NF02	As consultas deverão ser rápidas e seguras
NF03	A construção do <i>Data Warehouse</i> não deverá afetar o restante do funcionamento do portal

Fonte: O Autor (2018)

### 3.2.4 Data warehouse

Para atingir o sucesso deste projeto, será seguido o procedimento de construção de *Data Warehouse* seguindo alguns artifícios que foram descritos na Seção 2. Como o procedimento de Extração, Tratamento e Carregamento dos Dados (ETL) para o DW, para posteriormente estes dados serem visualizados por uma aplicação *front-end*, através de um *dashboard*. A Figura 23 representa a arquitetura de como ficará o DW.

Figura 23: Arquitetura de software



Fonte: O Autor (2018)

O banco de dados utilizado para construção do portal LACOS Web foi o PostgreSQL, e portanto para a construção do *Data Warehouse* será utilizado o mesmo banco de dados, apenas criando novas tabelas específicas para o sistema de BI proposto.

A forma como está sendo construído o sistema é similar a forma defendida por Kimball(2002), *bottom-up*, já que primeiramente está se projetando os baixos níveis, para posteriormente projetar os altos níveis, da mesma forma como será a construção deste sistema. Portanto, foram analisadas todas as tabelas e campos existentes no LACOS Web, identificando os componentes que seriam úteis ao BI. O Quadro 8 evidencia as tabelas do

LACOS Web e como serão utilizadas no *Data Warehouse*:

Quadro 8: Tabelas analisadas

(Continua)

<b>Tabela LACOS Web</b>	<b>Tabela <i>Data Warehouse</i></b>	<b>Tipo</b>	<b>Observação</b>
estado	d_estado	Dimensão	
município	d_município	Dimensão	
local_coleta	d_local_coleta	Dimensão	
tipo_coleta	d_tipo_coleta	Dimensão	
ponto_coleta	d_ponto_coleta	Dimensão	
coletores			Tabela não relevante para o BI
col_coletor			Tabela não relevante para o BI
coletas	f_coletas	Fato	Tabela fato, possuindo chave primária composta pelos campos id_coleta, data_coleta e data_ultchave. Os campos condmeteorologica, altantropicas, observacao e detalhes_vegetacao não serão relevantes ao BI. Os campos hora_coleta e fuso_horario serão tratados e inseridos junto ao campo data_coleta.
situacao_visual	d_situacao_visual	Dimensão	
profundidade_media	d_profundidade_media	Dimensão	
largura_media	d_largura_media	Dimensão	
velocidade_agua	d_velocidade_agua	Dimensão	
vegetacao_aquatica	d_vegetacao_aquatica	Dimensão	
tipos_fundo	d_tipos_fundo	Dimensão	
valor_medicao	d_valor_medicao	Dimensão	
tipos_medicao	d_tipos_medicao	Dimensão	
unidade_medida	d_unidade_medida	Dimensão	

(Conclusão)

<b>Tabela LACOS Web</b>	<b>Tabela Data Warehouse</b>	<b>Tipo</b>	<b>Observação</b>
unidade_medida	d_unidade_medida	Dimensão	
variaveis	d_variaveis	Dimensão	
categoria_profundidade	d_categoria_profundidade	Dimensão	
coletas_fauna	D_coletas_fauna	Dimensão	
dados_biologicos	f_dados_biologicos	Fato	Tabela fato, possuindo a mesma chave primária da tabela dados_biologicos. O campo secao não será relevante ao BI.
filo	d_dilo	Dimensão	
classe	d_classe	Dimensão	
subclasse	d_subclasse	Dimensão	
ordem	d_ordem	Dimensão	
família	d_familia	Dimensão	
genero	d_genero	Dimensão	
especie	d_especie	Dimensão	

Fonte: O Autor (2018)

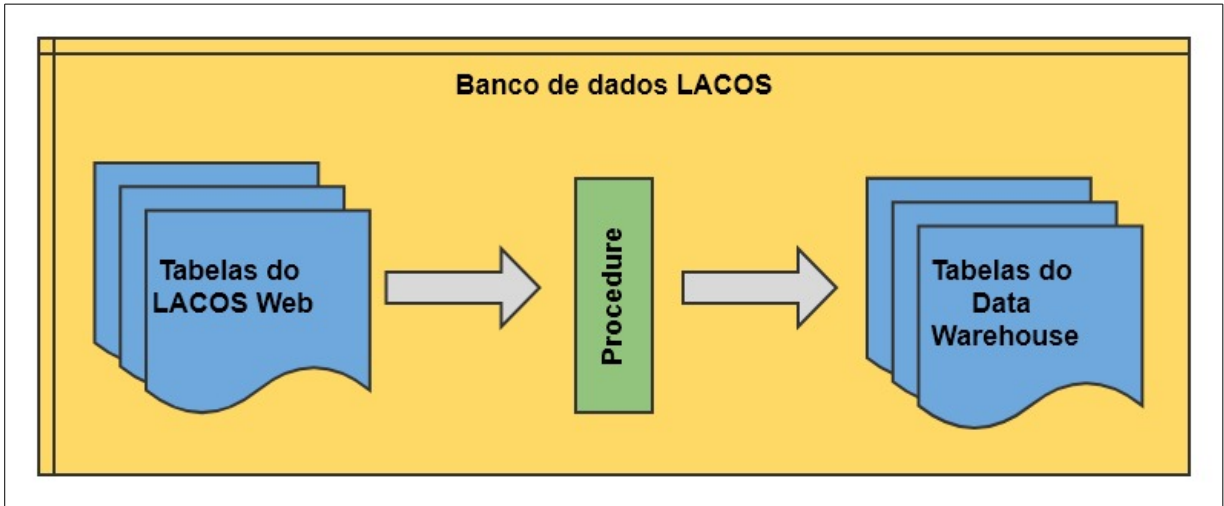
Por fins de organização, todas as tabelas fatos terão seus nomes iniciados pelo prefixo f\_ e todas as tabelas de dimensões pelo prefixo d\_. Neste modelo existem 3 tabelas fatos f\_coletas, f\_valor\_medicao e f\_dados\_biologicos, e todas as dimensões ao redor destas tabelas hierarquizadas compondo um esquema floco de neve. Todas as tabelas de dimensão deverão ter os mesmos registros que a sua tabela equivalente no LACOS Web. O modelo dimensional foi construído embasado no modelo lógico atual, o qual as tabelas dimensão que irão compor o DW possuem estrutura idêntica as tabelas do LACOS Web. Sendo assim o modelo lógico pode ser considerado parcialmente dimensional, o que auxiliou na elaboração do modelo elaborado e que está exposto na Figura 24.





Foi desenvolvido uma procedure no PostgreSQL a qual realiza o procedimento de ETL dos dados a partir das tabelas do LACOS Web para as tabelas do *Data Warehouse*, conforme representado pela Figura 25.

Figura 25: Procedimento de ETL para o LACOS Web



Fonte: O Autor (2018)

Esta procedure que realiza o ETL está descrita conforme o algoritmo exposto no Quadro 9, em linguagem figurativa, a qual foi codificado em PL/pgSQL<sup>20</sup>:

Quadro 9: Processo de ETL

<b>Procedure</b>
Para cada tabela do LACOS Web que existe uma tabela dimensão no <i>Data Warehouse</i> , insira todos os seus registros na tabela do <i>Data Warehouse</i> equivalente.
Para cada registro da tabela <i>valor_medicao</i> , trate esse registro e insira na tabela <i>f_valor_medicao</i> .
Para cada registro da tabela <i>coletas</i> , trate esse registro e insira na tabela <i>f_coletas</i> .
Para cada registro da tabela <i>dados_biologicos</i> , trate esse registro e insira na tabela <i>f_dados_biologicos</i> .
Qualquer exceção que ocorrer, o registro será descartado.
Manda aviso ao usuário com a quantidade de registros que foram importados e que ocorrem algum erro.

Fonte: O Autor (2018)

<sup>20</sup> PL/pgSQL: linguagem estrutural estendida da SQL que tem por objetivo auxiliar as tarefas de programação no PostgreSQL

### 3.2.5 INTERFACE GRÁFICA

Exemplos de métricas (Quadro 10) foram apresentadas aos pesquisadores e sendo amadurecidas durante a implementação.

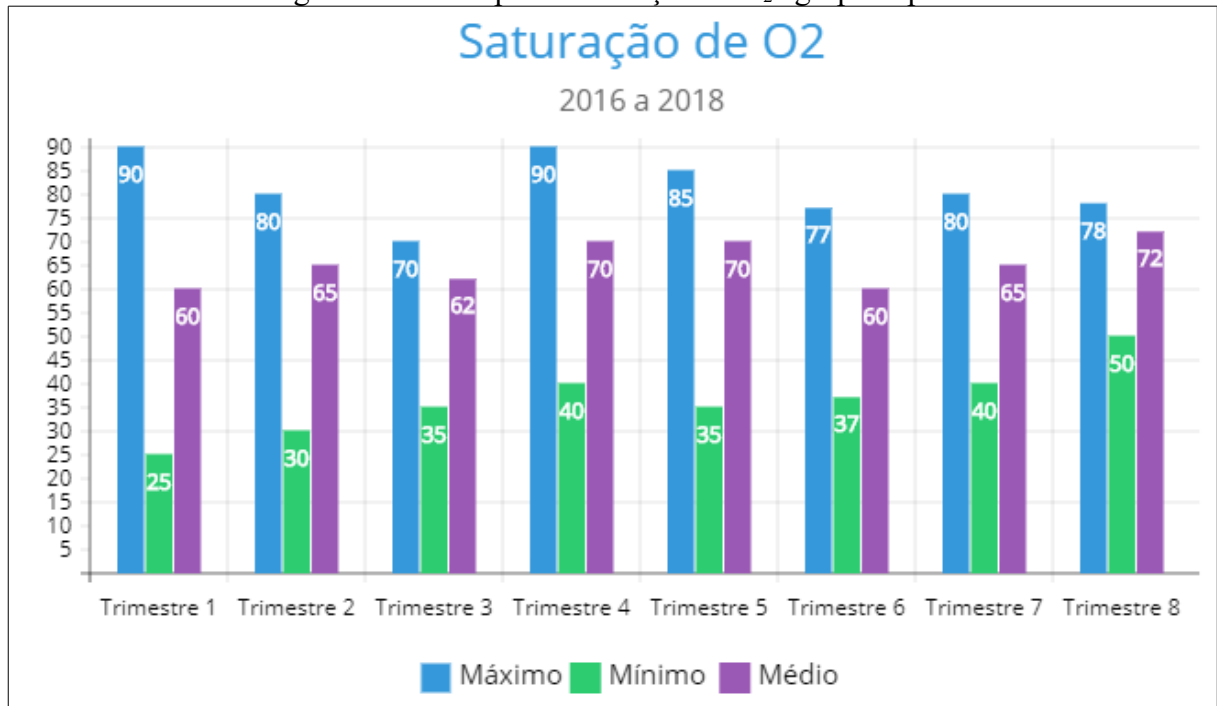
Quadro 10: Exemplo de métricas

Identificador	Descrição
MT01	Permitir visualizar valor máximo, mínimo e médio de PH da água
MT02	Permitir visualizar valor máximo, mínimo e médio de saturação de O <sub>2</sub> de água
MT03	Permitir visualizar temperatura máxima, mínima e média
MT04	Permitir visualizar velocidade máxima, mínima e média do vento
MT05	Permitir visualizar média de transparência da água
MT06	Permitir visualizar profundidade mínima, máxima e média

Fonte: O Autor (2018)

O protótipo da Figura 26, exemplifica um possível gráfico de saturação de O<sub>2</sub>, no período de 2016 a 2018, agrupando suas métricas (máximo, mínimo e médio) por trimestre.

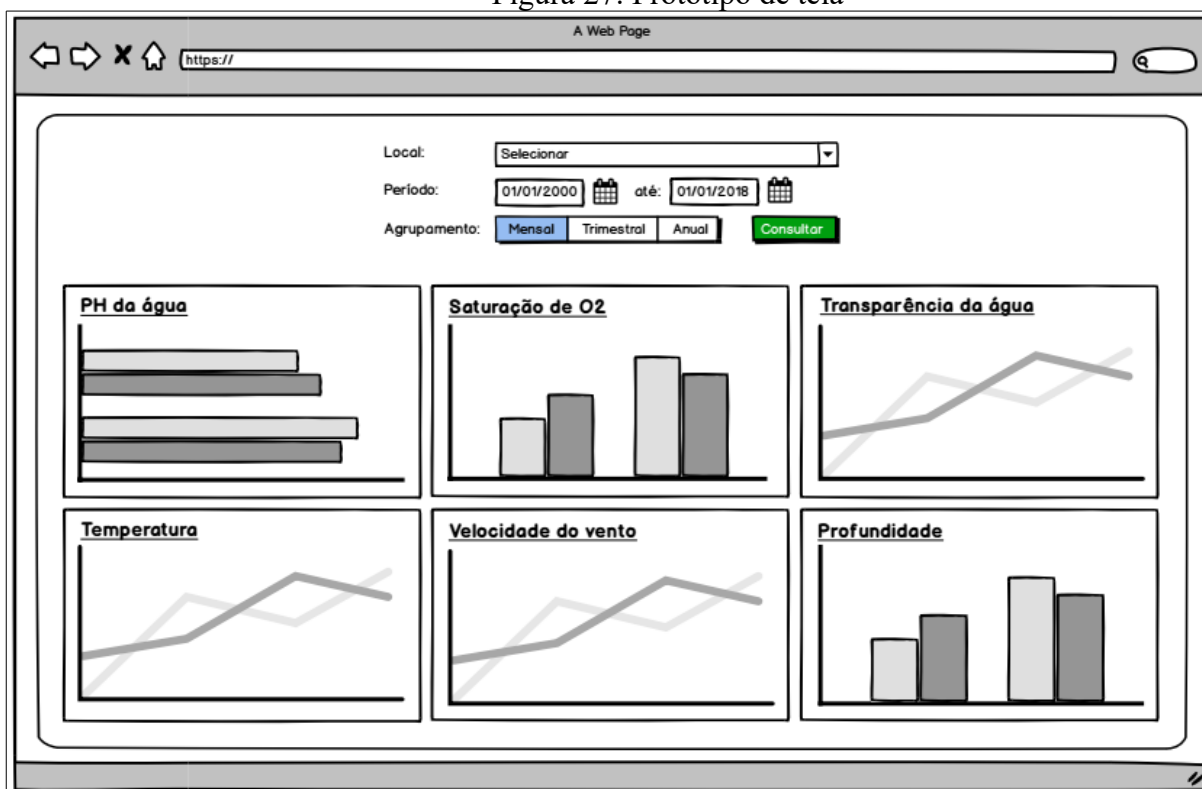
Figura 26: Protótipo de saturação de O<sub>2</sub> agrupado por trimestre



Fonte: O Autor (2018)

A camada de apresentação foi projetada dentro dos componentes que a linguagem PHP proporciona. Foram construídas consultas que exibirão gráficos de barras horizontais e verticais, e gráficos de linhas, conforme o protótipo da Figura 27 que possui as métricas estipuladas pelo Quadro 10, além de outras formas que foram propostas pelos pesquisadores.

Figura 27: Protótipo de tela



Fonte: O Autor (2018)

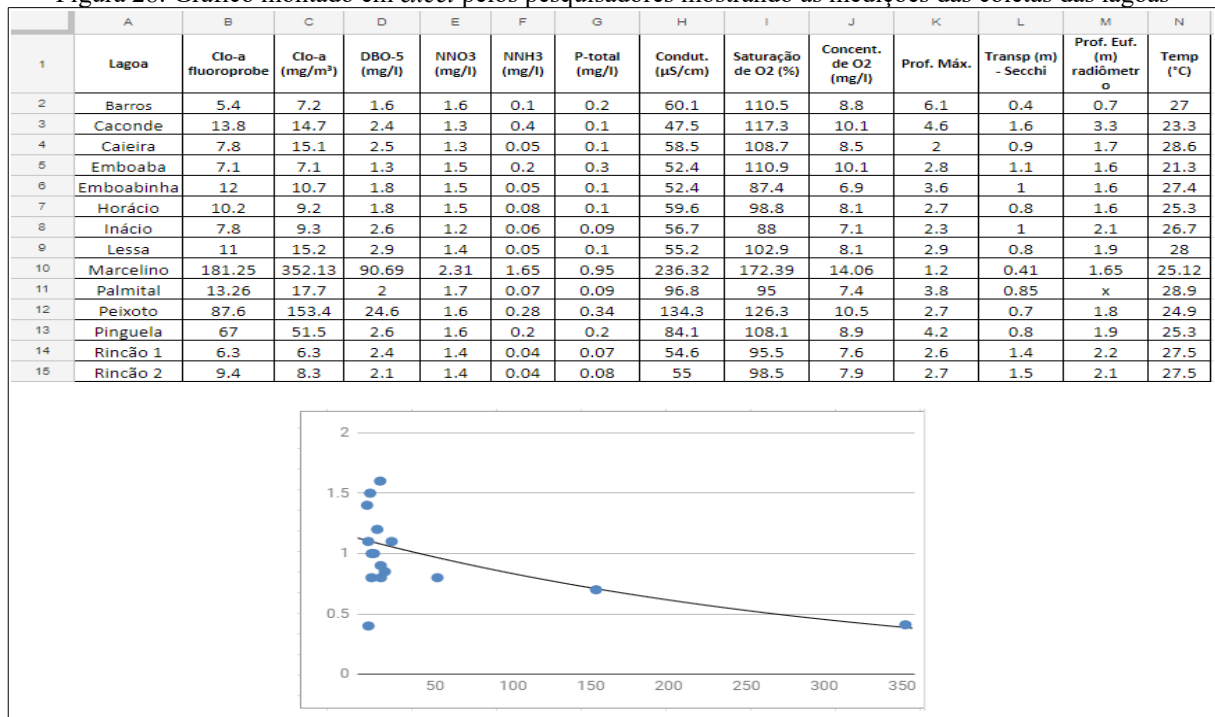
## 4 IMPLEMENTAÇÃO

A implementação do módulo de *Business Intelligence* ao portal LACOS Web foi construída incrementalmente, desde a criação de tabelas que compõem o *Data Warehouse* até os gráficos gerados na camada de apresentação. Neste capítulo será apresentado o desenvolvido deste novo módulo, elucidando cada passo até o modelo proposto ser construído.

### 4.1 SITUAÇÃO ATUAL

Inicialmente foram realizadas reuniões com os pesquisadores do Projeto LACOS. Como resultado destas reuniões, definiu-se as formas de cruzamento de dados e quais gráficos serão utilizados. Atualmente, os gráficos são criados em planilhas excel ou em outras ferramentas, onde cabe ao pesquisador fornecer todas as medidas e gerar manualmente o gráfico desejado. A Figura 28 apresenta uma série de coletas e um gráfico gerado pelo usuário em uma planilha *Excel*.

Figura 28: Gráfico montado em *excel* pelos pesquisadores mostrando as medições das coletas das lagoas

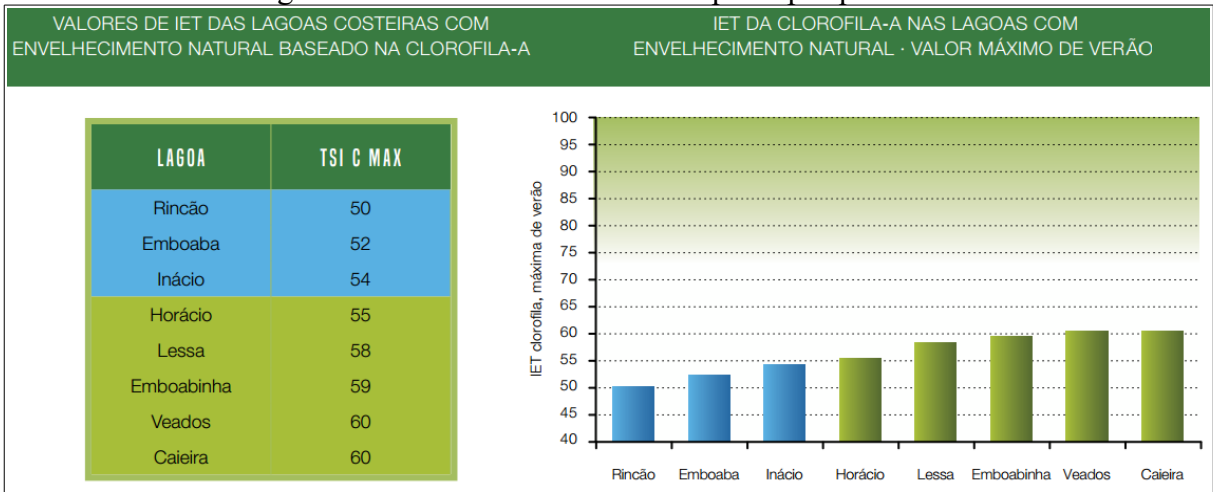


Fonte: LACOS (2018)

Outro meio utilizado pelos pesquisadores para apresentar informações é gerando o gráfico através de alguma ferramenta específica, tal como foi feito para exibir as informações no Atlas Socioambiental do Município de Osório<sup>21</sup>, apresentadas nas Figuras 29 e 30.

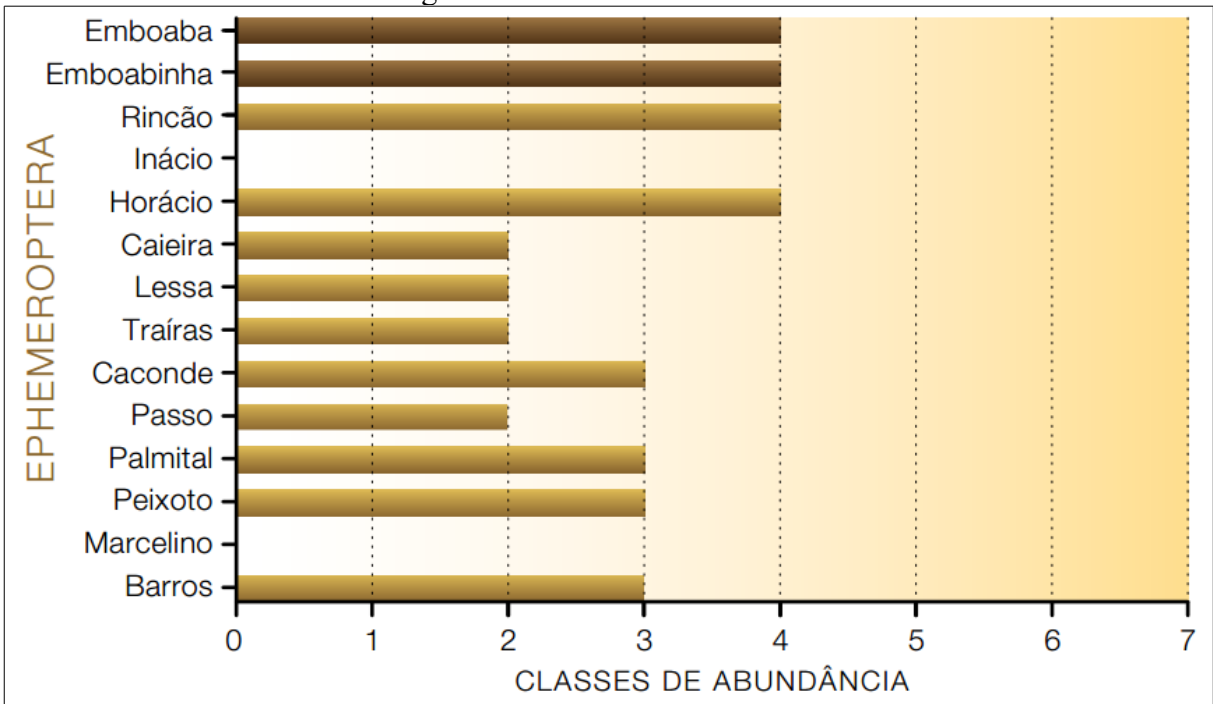
21 Atlas Socioambiental do Município de Osório: Disponível em: <https://www.ucs.br/site/midia/arquivos/ebook-atlas-projeto-lacos.pdf>. Acessado em 13/05/2018

Figura 29: Gráficos atuais utilizado pelos pesquisadores



Fonte: SCHÄFER (2013)

Figura 30: Gráfico atual utilizado



Fonte: SCHÄFER (2013)

Com exemplares destes gráficos adquiridos e reuniões junto aos envolvidos realizadas a fim de levantar os requisitos e as necessidades dos pesquisadores, deu-se início a fase de análise destes materiais, onde conseguiu-se compreender como seria possível atingir os objetivos deste trabalho.

#### 4.2 CONSTRUÇÃO DO DATA WAREHOUSE

Com base na modelagem do banco de dados do portal LACOS Web foi proposto o modelo multidimensional, conforme apresentado na Seção 3.2.4. A partir deste foram gerados os *scripts* de criação de tabelas do *Data Warehouse*. A Figura 31 apresenta o código para

criação da tabelas dimensão d\_unidade\_medida, d\_variaveis, d\_tipo\_medicao, já sendo populadas com os dados das tabelas do banco de dados do LACOS Web.

Figura 31: Script de criação das tabela d\_unidade\_medida, d\_variaveis e d\_tipo\_medicao

```
create table d_unidade_medida as select * from unidade_medida;
alter table d_unidade_medida add constraint
d_unidade_medida_pkey PRIMARY KEY (id_unimedida);

create table d_variaveis as select * from variaveis;
alter table d_variaveis add constraint
d_variaveis_pkey PRIMARY KEY (id_variavel);

create table d_tipo_medicao as select * from tipo_medicao;
alter table d_tipo_medicao add constraint
d_tipo_medicao_pkey PRIMARY KEY (id_tpmedicao);
```

Fonte: O Autor (2018)

E a Figura 32 apresenta um trecho do código da criação da tabela fato f\_valor\_medicao.

Figura 32: Script de criação da tabela fato f\_valor\_medicao

```
create table f_valor_medicao(
id_valormedicao bigint not null
, id_coleta integer
, id_variavel integer
, id_tpmedicao integer
, id_unimedida integer
, id_categoriaprof integer|
, medicao numeric(10,2)
, valor numeric(10,5));

alter table f_valor_medicao
add constraint f_valor_medicao_pkey PRIMARY KEY (id_valormedicao);

ALTER TABLE f_valor_medicao
ADD CONSTRAINT f_valor_medicao_id_tpmedicao_fkey FOREIGN KEY (id_tpmedicao)
REFERENCES d_tipo_medicao (id_tpmedicao);

ALTER TABLE f_valor_medicao
ADD CONSTRAINT f_valor_medicao_id_variavel_fkey FOREIGN KEY (id_variavel)
REFERENCES d_variaveis (id_variavel);
```

Fonte: O Autor (2018)

E para finalizar a construção do *DW* e procedimento de ETL, foi desenvolvido uma *procedure* a qual realiza a carga das tabelas fato e dimensão como havia sido previamente definido. A Figura 33 apresenta um trecho da *procedure* INSERE\_DW, através da qual são inseridos dados na tabela fato f\_coletas\_fauna.

Figura 33: Procedure INSERE\_DW

```

CREATE OR REPLACE FUNCTION INSERE_DW ()
RETURNS void AS '
BEGIN
...
INSERT INTO F_COLETAS_FAUNA
        (id_coleta
        ,data_coleta
        ,substrato
        ,id_ponto
        ,tipo_coleta)
SELECT CF.id_coleta
        ,CF.data_coleta
        ,CF.substrato
        ,CF.id_ponto
        ,CF.tipo_coleta
FROM COLETAS_FAUNA CF
WHERE NOT EXISTS (SELECT 1
                   FROM F_COLETAS_FAUNA FCF
                   WHERE FCF.id_coleta = CF.id_coleta);
...
END;'
LANGUAGE 'plpgsql';

```

Fonte: O Autor (2018)

#### 4.3 CAMADA DE APRESENTAÇÃO

Seguindo o padrão de desenvolvimento do portal LACOS Web, para a construção deste módulo de BI adotou-se a linguagem de programação PHP, acompanhado dos demais *frameworks* que já vinham sendo utilizados, tais como *CodeIgniter*<sup>22</sup> e *Bootstrap*. Para a exibição dos gráficos, entretanto, foi utilizado a biblioteca *Google Charts*, a qual consegue suprir a necessidades dos usuários por possuir um grande leque de possibilidades de exibição gráficas e ser uma biblioteca altamente difundida, facilitando assim a adaptação dos pesquisadores. A biblioteca *D3js*<sup>23</sup> também foi avaliada, porém decidiu-se não utilizar por possuir maior complexidade, e ser menos difundida no mercado atual.

Conforme conversado em reunião com os pesquisadores, os gráficos que preferencialmente serão utilizados são pizza, colunas, barras e linhas, a fim de demonstrar os cruzamentos de variáveis.

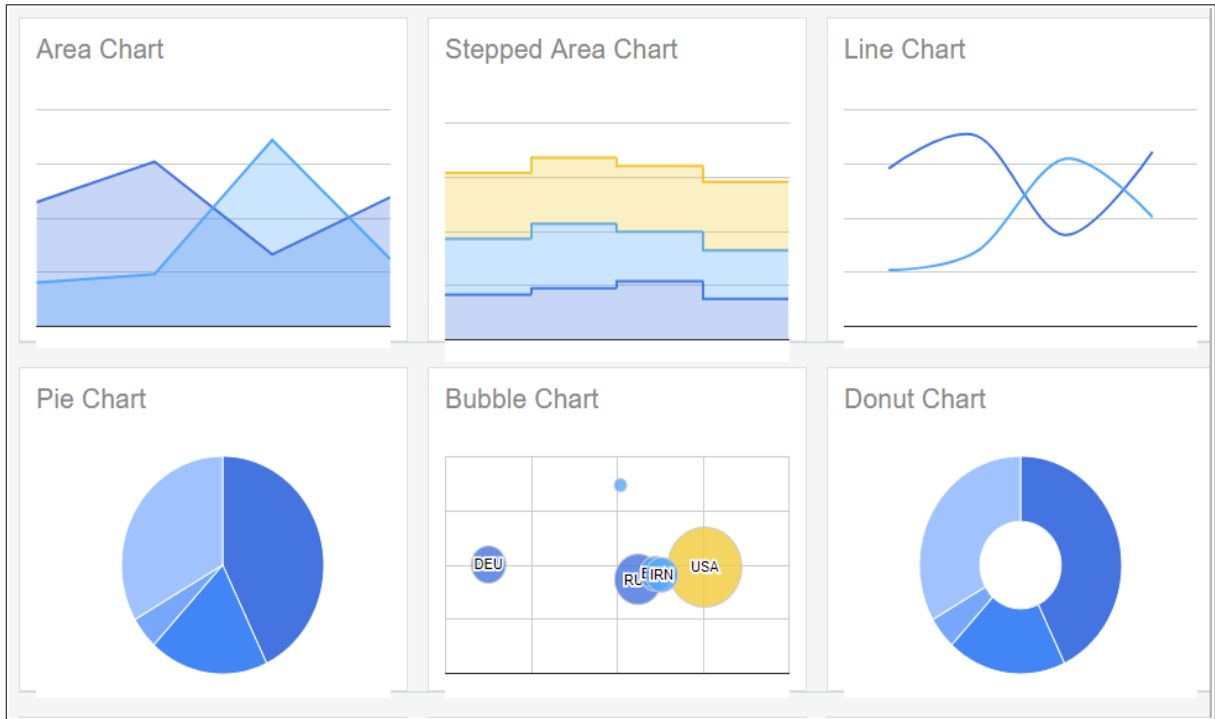
22 *CodeIgniter: Framework* de desenvolvimento em PHP. Disponível em: <https://codeigniter.com/> acessado em 14/05/2018.

23 *D3js*: <https://d3js.org/> acessado em 25/06/2018.



A Figura 34 exibe algumas formas de gráficos disponíveis pela biblioteca *Google Charts*.

Figura 34: Formas de gráficos da biblioteca Google Charts



Fonte: Google (2018)

#### 4.4 CONSIDERAÇÕES FINAIS

Dado o momento que foi disponibilizado o acesso ao banco de dados do LACOS, iniciou-se a construção das tabelas do *Data Warehouse* conforme o modelo proposto. Tendo em vista o ganho em performance da rotina de ETL, os *scripts* de criação de cada tabela dimensão do DW foi feita com a instrução de criação e inserção dos dados em questão. Também pensando em performance, a rotina de ETL foi desenvolvida agindo somente em dados que ainda não estejam no DW, mesmo assim simulações de erros na inserção dos dados foram realizadas, a fim de identificar a exibição destes erros na aplicação.

A partir do momento em que foi realizado o ETL, buscou-se compreender as formas de gráficos desejadas pelos usuários, as quais foram concebidas através das reuniões realizadas, e que a cada reunião foram otimizadas conforme apresentadas à pesquisadora. Portanto, os *dashboards* foram sendo adaptados ao longo do projeto até atingirem a necessidade dos pesquisadores.

## 5 ESTUDO DE CASO

Este capítulo apresentará as novas funcionalidades implementadas ao portal LACOS Web, as consultas realizadas através de cruzamentos de informações e as formas de exibição dos dados do *Data Warehouse*, sintetizadas através de gráficos que compõem o módulo de BI.

### 5.1 MÓDULO DE BI

O acesso ao módulo de BI, bem como a qualquer outro módulo do portal LACOS Web, exige o usuário ter permissão de acesso. E é esta a única premissa para o usuário ter acesso as funcionalidades do módulo, o qual está dividido em 3 submódulos:

- **BI – Físico/Químicos:** onde estão disponíveis informações cruzadas a cerca de dados físicos e químicos
- **BI – Fauna:** onde estão disponíveis informações cruzadas a cerca de dados de fauna
- **BI – Administrativo:** onde está localizado a parte de ETL para o *Data Warehouse*.

#### 5.1.1 BI – FÍSICO/QUÍMICO

O BI de dados físicos e químicos é apresentado na Figura 35, e possui uma série de filtros que devem ser informadas pelo usuário, a fim de localizar as informações e imprimi-las na tela.

Figura 35: Filtros do BI – Físico/Químico

The screenshot displays the 'Business Intelligence - Físico/Químico' interface. On the left is a dark sidebar with navigation options: Home, Administrativo, Dados Físico/Químicos (selected), Dados Biológicos, Georreferenciamento, Exportação, and BI. The main content area is titled 'Business Intelligence - Físico/Químico' and contains a 'Filtros' section with the following elements:

- Data Inicial:** Input field with placeholder 'dd/mm/aaaa'.
- Data Final:** Input field with placeholder 'dd/mm/aaaa'.
- Variável:** Dropdown menu with 'Alcalinidade' selected.
- Local:** Radio buttons for 'Todos' (selected) and 'Específico'.
- Tipo de Local:** Dropdown menu with 'Todos' selected.
- Local da Coleta:** Dropdown menu with 'Arroio Barrinha' selected.
- Gráfico:** Radio buttons for 'Linhas (Profundidade no eixo X)', 'Linhas (Profundidade no eixo Y)', and 'Barras'.
- Desvio padrão**

At the bottom of the filter section are two buttons: 'Consultar' (green) and 'Limpar' (red).

Fonte: O Autor(2018)

Os filtros de busca apresentados na Figura 35 estão explanados através do Quadro 11.

Quadro 11: Filtros do BI – Físico/Químico

<b>Filtro</b>	<b>Observação</b>
<b>Data inicial</b>	Período inicial da data de coleta. Esta opção é opcional.
<b>Data final</b>	Período final da data de coleta. Esta opção é opcional.
<b>Variável</b>	Permite ao usuário selecionar uma variável de coleta.
<b>Local</b>	Duas opções são dadas ao usuário: <b>1. Todos:</b> onde será exibido um gráfico contendo todos os locais onde foram coletadas a variável de coleta selecionada. <b>2. Específico:</b> permite ao usuário selecionar um local específico de coleta.
<b>Tipo de Local</b>	Opção ativa somente quando o Local selecionado for Todos, possibilitando ao usuário selecionar o tipo de local onde foram feitas as coletas.
<b>Local da Coleta</b>	Opção ativa somente quando o Local selecionado for Específico e dá a possibilidade do usuário selecionar o local da coleta.
<b>Gráfico</b>	Existem três opções as quais indicam como será plotado o gráfico: <b>1. Linhas (Profundidade no eixo X):</b> onde será impresso o valor das coletas em forma de linhas no eixo Y e as profundidades no eixo X. <b>2. Linhas (Profundidade no eixo Y):</b> onde será impresso o valor das profundidades em forma de linha no eixo Y e o valor das coletas no eixo X. <b>3. Barras:</b> imprimirá o gráfico em forma de barras verticais.
<b>Desvio Padrão</b>	Permite ao usuário marcar se deseja exibir no gráfico os desvios padrão das medidas apresentadas. Opção disponível para Gráficos em Linhas (Profundidade no eixo X) e Barras.

Fonte: O Autor (2018)

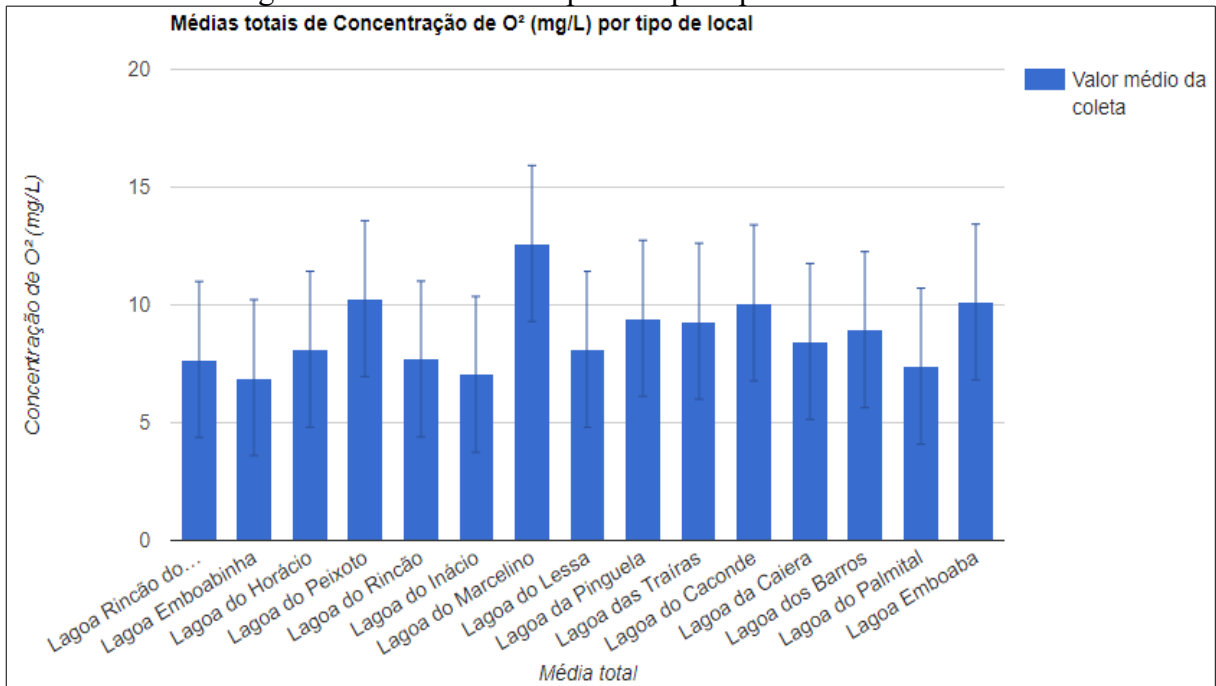
Estes critérios de busca conseguem ser combinados entre si e exibem de diferentes formas os gráficos ao usuário. Tais formas de resultado são explicadas na Seção 5.1.1.1 a Seção 5.1.1.4.

#### 5.1.1.1 Gráfico físico/químico contendo valores médios de coletas por tipos de local

Para o usuário imprimir este tipo de gráfico é necessário selecionar a opção de Todos os locais, podendo também escolher algum tipo de local. O resultado será a impressão da média de valores de coletas por local. Por exemplo, a Figura 36 é o resultado dos seguintes filtros:

- **Variável:** Concentração de O<sup>2</sup> (mg/L)
- **Local:** Todos
- **Tipo de local:** Lagoa
- **Desvio padrão:** X
- **Demais filtros de busca não foram informados**

Figura 36: Gráfico físico/químico por tipo de local - Barras



Fonte: O Autor (2018)

Percebe-se que são impressas barras verticais e que no topo de cada barra encontra-se uma pequena linha, também na vertical, a qual vai para a direção superior e também para a inferior, isso para representar o desvio padrão do conjunto de dados selecionados.

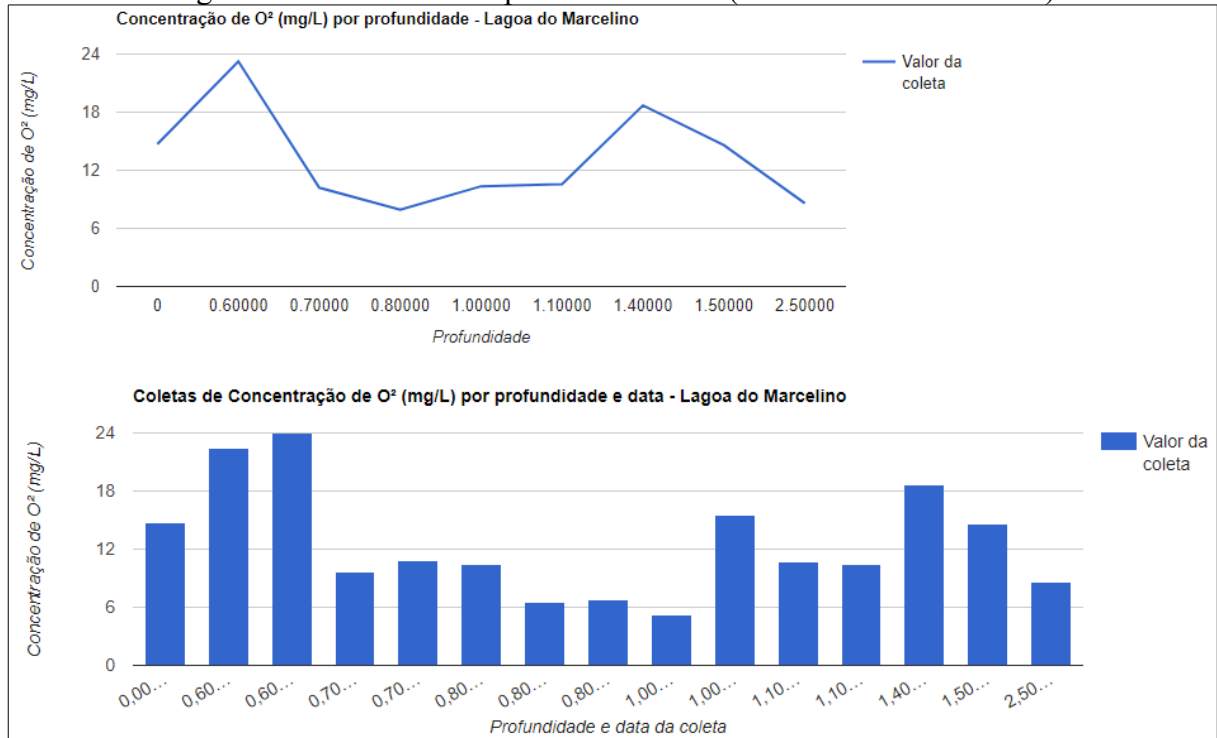
#### 5.1.1.2 Gráfico biológico contendo valores médios de coletas por profundidade do local – Linhas (Profundidade no eixo X)

Para a impressão deste gráfico é necessário selecionar a opção de Local Especifico, algum Local de coleta e o Gráfico em Linhas (Profundidade no eixo X). Na Figura 37 é apresentado uma impressão deste modelo de gráfico, o qual foi gerado a partir das seguintes opções:

- **Variável:** Concentração de O<sup>2</sup> (mg/L)
- **Local:** Específico
- **Local da coleta:** Lagoa do Marcelino
- **Gráfico:** Linhas (profundidade no eixo X)
- **Demais filtros de busca não foram informados**

A Figura 37 exibe dois gráficos, sendo que no primeiro deles o valor da coleta está agrupado por profundidade, sendo calculado pela média aritmética de todas as coletas da profundidade em questão. O segundo gráfico, em barras, mostra todas as coletas realizadas exibindo também a data da coleta e a profundidade.

Figura 37: Gráfico físico/químico – Linhas (Profundidade no eixo X)



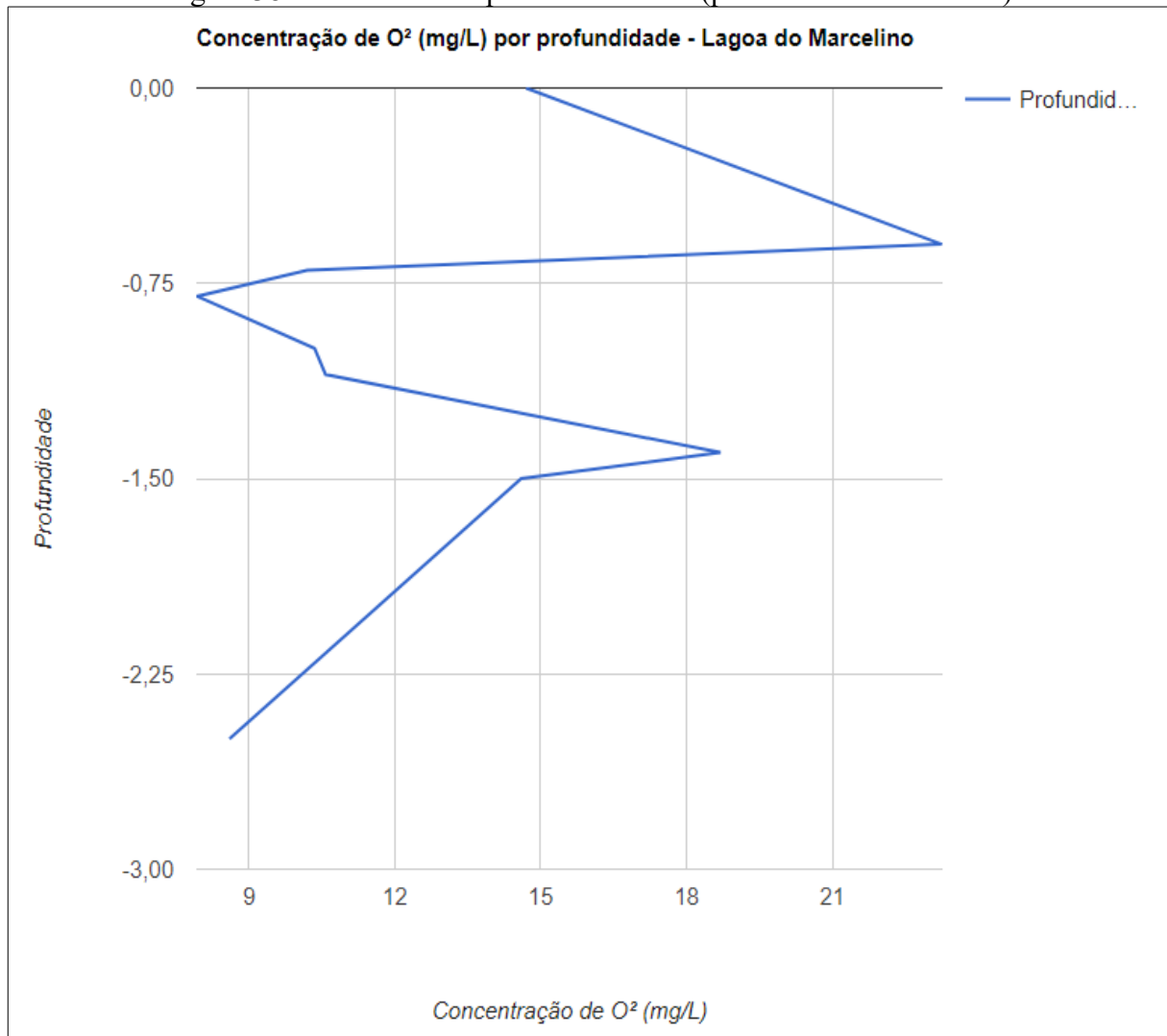
Fonte: O Autor (2018)

#### 5.1.1.3 Gráfico físico/químico contendo valores médios de coletas por profundidade do local – Linhas (Profundidade no eixo Y)

Este *dashboard* apresenta as mesmas informações dos gráficos citados na Seção 5.1.1.2, com a diferença que a forma de apresentar o gráfico está rotacionada em 90° para a direita, mostrando no eixo Y as profundidades do local da coleta e no eixo X a média das coletas naquela profundidade. E o segundo gráfico, é apresentado da mesma maneira, listando todas as coletas realizadas e suas datas. A Figura 38 exemplifica uma execução com os seguintes filtros:

- **Variável:** Concentração de O<sup>2</sup> (mg/L)
- **Local:** Específico
- **Local da coleta:** Lagoa do Marcelino
- **Gráfico:** Linhas (profundidade no eixo Y)
- **Demais filtros de busca não foram informados**

Figura 38: Gráfico físico/químico – Linhas (profundidade no eixo Y)



Fonte: O Autor (2018)

Nota-se que o resultado é exatamente o mesmo da Seção 5.1.1.2, porém o gráfico apresenta as informações nos eixos contrários.

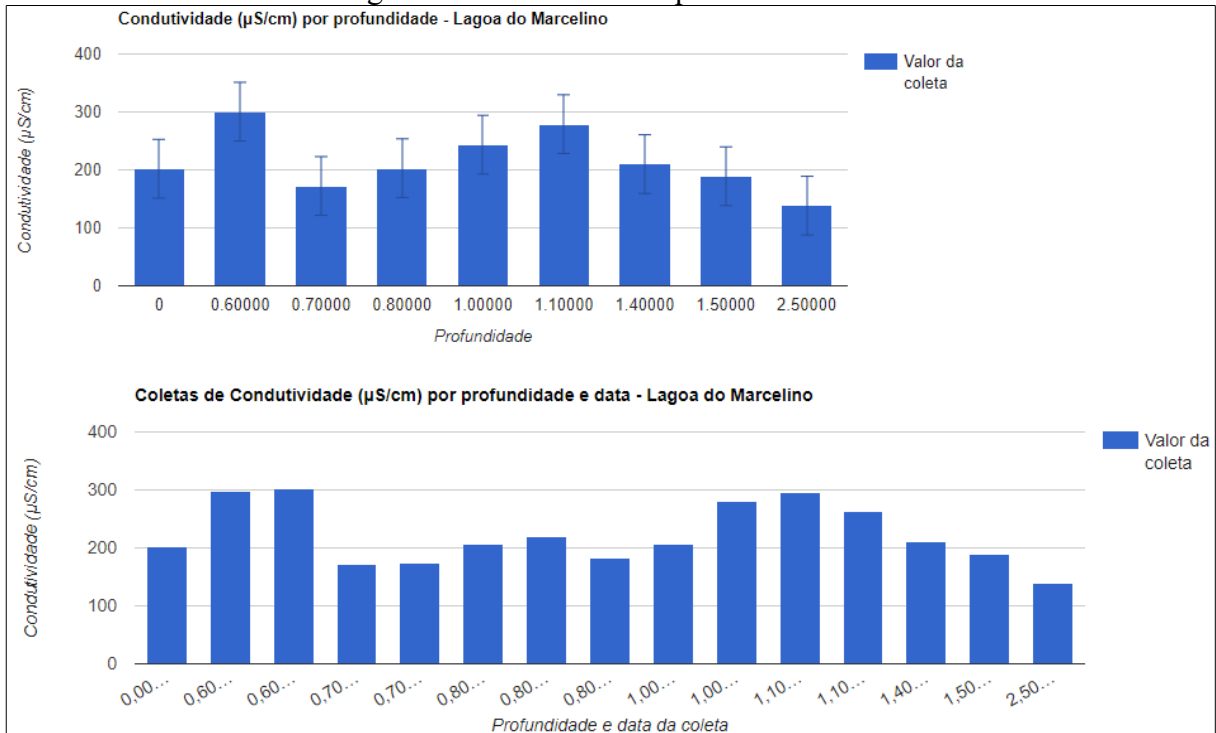
#### 5.1.1.4 Gráfico físico/químico contendo valores médios de coletas por profundidade do local – Barras

Também foi desenvolvido uma forma de imprimir o gráfico biológico em forma de barras, o qual é apresentado na Figura 39 e para plotar este gráfico foram utilizados os seguintes filtros:

- **Variável:** Condutividade ( $\mu\text{S}/\text{cm}$ )
- **Local:** Específico
- **Local da coleta:** Lagoa do Marcelino
- **Gráfico:** Barras

- Desvio padrão: X
- Demais filtros de busca não foram informados

Figura 39: Gráfico físico/químico - Barras



Fonte: O Autor (2018)

Neste *dashboard* são exibidos dois gráficos, sendo que no primeiro deles as informações estão agrupadas por profundidade e o valor apresentado na barra é o valor médio de todas as coletas realizadas naquela profundidade. Já no segundo gráfico, assim como apresentado na Seção 5.1.1.2, são exibidas todas as coletas realizadas por profundidade, local e data.

### 5.1.2 BI – FAUNA

O módulo de BI de fauna foi construído para permitir aos pesquisadores visualizar informações a cerca da presença de animais por locais. Permite ao usuário imprimir gráficos que expõem quantitativamente a presença de certos animais por lagoa ou as lagoas onde existem determinadas espécies de animais. A Figura 40 apresenta a interface gráfica inicial deste módulo.

Figura 40: Filtro do BI - Fauna

Fonte: O Autor (2018)

Os filtros de pesquisa existente para o módulo de Fauna são os descritos no Quadro 12.

Quadro 12: Filtros do BI – Fauna

(Continua)

Filtro	Observação
<b>Data inicial</b>	Período inicial da data de coleta. Esta opção é opcional.
<b>Data final</b>	Período final da data de coleta. Esta opção é opcional.
<b>Agrupamento</b>	Forma de agrupamento dos dados, sendo que existem duas opções ao usuário: <b>1. Local:</b> nesta opção os dados são agrupados por Local, sendo que é obrigatório o usuário escolher um local de coleta. <b>2. Espécie:</b> nesta opção os dados são agrupados por Espécie, e o usuário pode escolher o Filo, Classe, Subclasse, Ordem, Família, Gênero e Espécie.
<b>Local da coleta</b>	Opção ativa somente quando o Agrupamento selecionado for por Local e possibilita o usuário selecionar o local da coleta.
<b>Filo</b>	Opção ativa somente quando o Agrupamento selecionado for Espécie e possibilita o usuário selecionar o Filo.



(Conclusão)

<b>Filtro</b>	<b>Observação</b>
<b>Classe</b>	Opção ativa somente quando o Agrupamento selecionado for Espécie e possibilita o usuário selecionar a Classe.
<b>Subclasse</b>	Opção ativa somente quando o Agrupamento selecionado for Espécie e possibilita o usuário selecionar a Subclasse.
<b>Ordem</b>	Opção ativa somente quando o Agrupamento selecionado for Espécie e possibilita o usuário selecionar a Ordem.
<b>Família</b>	Opção ativa somente quando o Agrupamento selecionado for Espécie e possibilita o usuário selecionar a Família.
<b>Gênero</b>	Opção ativa somente quando o Agrupamento selecionado for Espécie e possibilita o usuário selecionar o Gênero.
<b>Espécie</b>	Opção ativa somente quando o Agrupamento selecionado for Espécie e possibilita o usuário selecionar a Espécie.
<b>Classificação</b>	Opção ativa somente quando o Agrupamento selecionado for Espécie e permite aos usuários duas possibilidades de seleção: <b>1. N° de indivíduos:</b> nesta opção será mostrado o número de indivíduos que efetivamente foram coletados. <b>2. Classe de abundância:</b> nesta opção o número de indivíduos é apresentado conforme a classe de abundância.
<b>Tipo de local</b>	Opção ativa somente quando o Agrupamento selecionado for Espécie e permite selecionar algum Tipo de local.

Fonte: O Autor (2018)

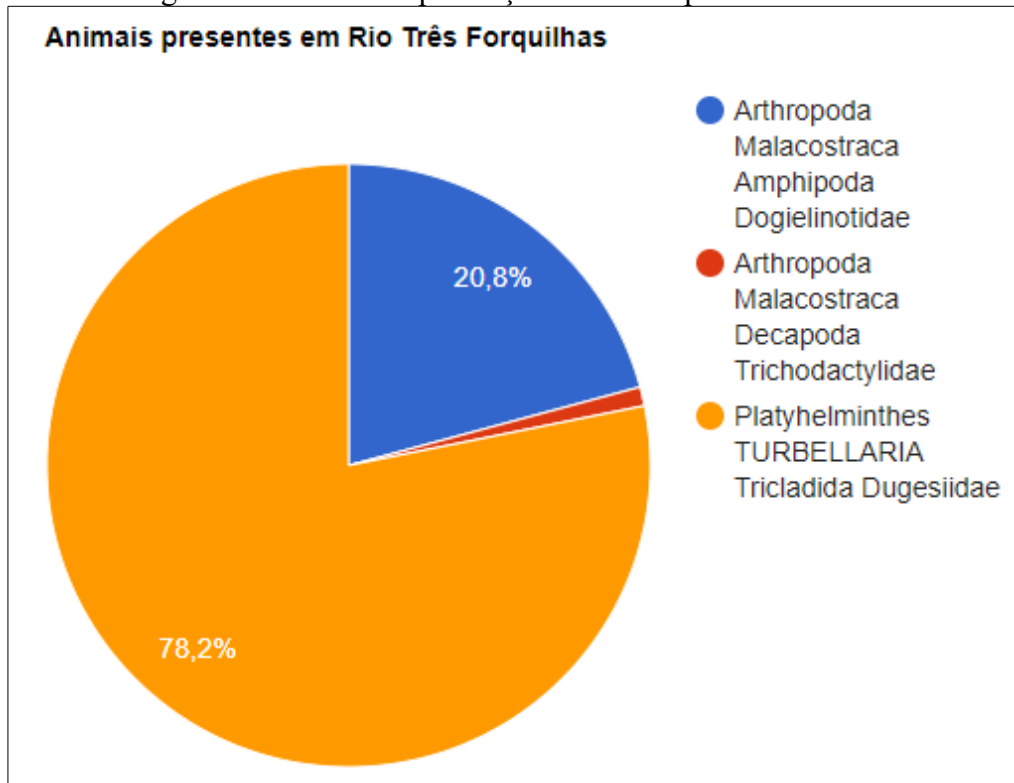
Estes filtros são capazes de gerar duas combinações de resultados possíveis, os quais serão explicados na Seção 5.1.2.1 e na Seção 5.1.2.2.

#### 5.1.2.1 Gráfico da presença de animais por local

Esta forma de gráfico permite ao usuário identificar a quantidade de animais em um local específico. Os seguintes filtros de busca foram utilizados para gerar um exemplo de execução e o resultado é apresentado através da Figura 41:

- **Data inicial:** 10/05/2005
- **Data final:** 20/02/2007
- **Agrupamento:** Local
- **Local da coleta:** Rio Três Forquilhas
- **Demais filtros de busca não foram informados**

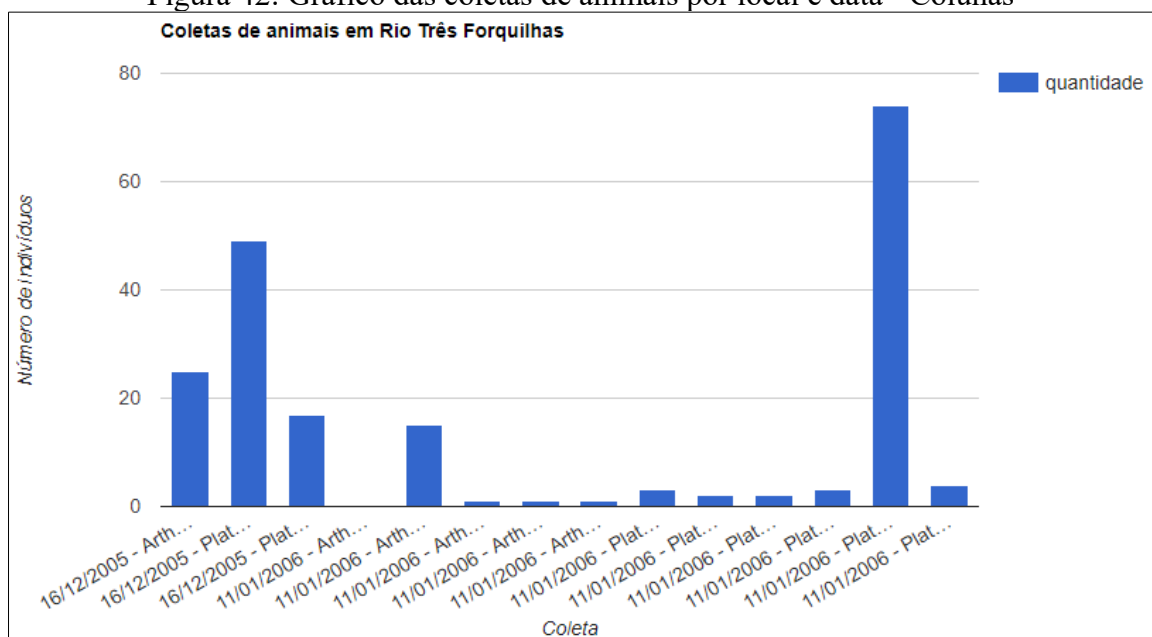
Figura 41: Gráfico da presença de animais por local - Pizza



Fonte: O Autor (2018)

Logo abaixo deste gráfico, assim como no BI Biológico, existe um gráfico em barras contendo todas as coletas realizadas e as suas respectivas datas de coleta, representado através da Figura 42.

Figura 42: Gráfico das coletas de animais por local e data - Colunas



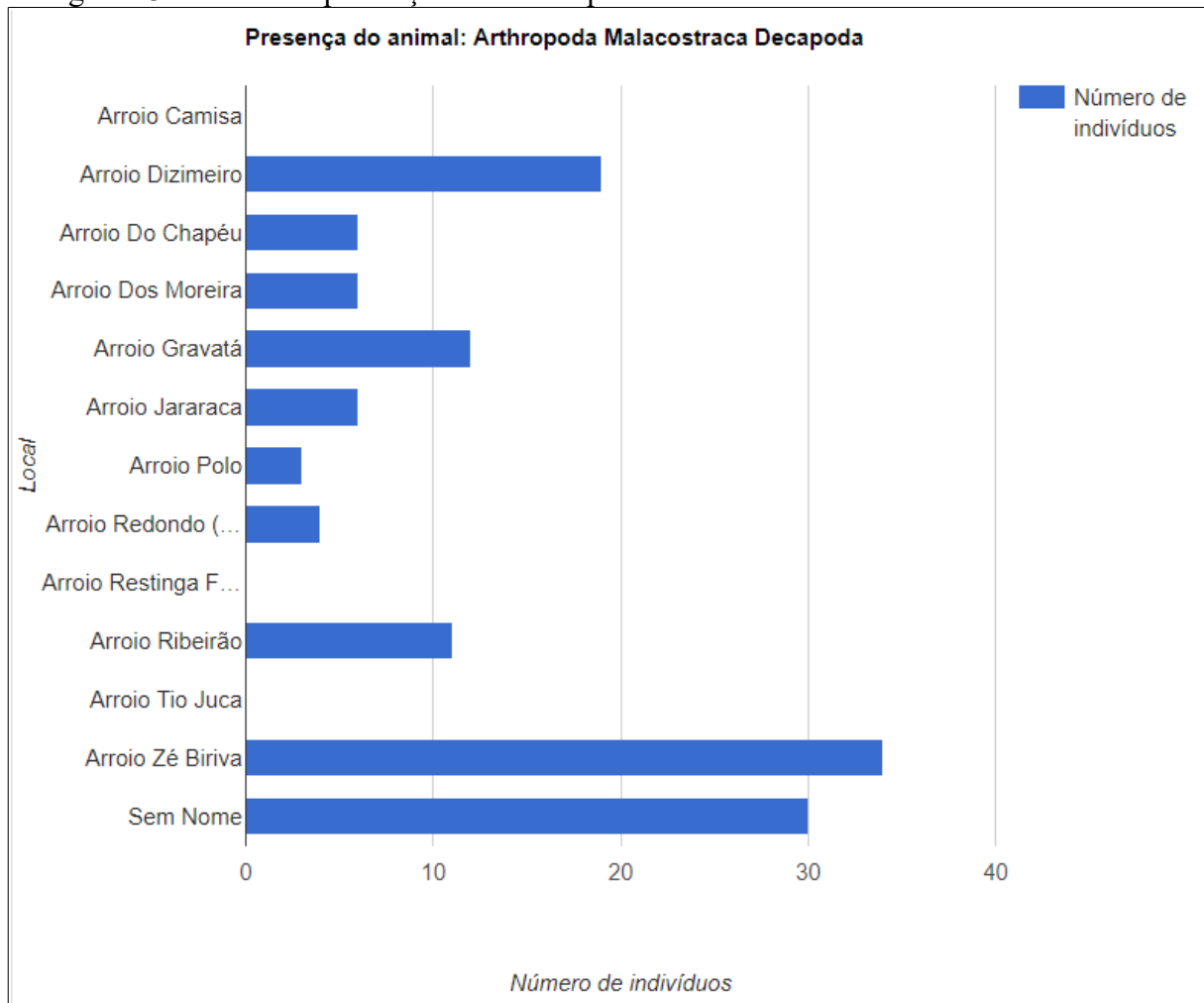
Fonte: O Autor (2018)

### 5.1.2.2 Gráfico da presença de espécies por locais

A segunda métrica que o módulo BI – Fauna provê é um gráfico onde exhibe as lagoas nas quais determinada espécie foi coletada. Por exemplo, o uso dos seguintes filtros de busca resultará no gráfico exposto pela Figura 43:

- **Agrupamento:** Espécie
- **Filo:** Arthropoda
- **Classe:** Malacostraca
- **Ordem:** Decapoda
- **Classificação:** N° de indivíduos
- **Tipo de local:** Arroio
- **Demais filtros de busca não foram informados**

Figura 43: Gráfico da presença de animais por locais – Barras com número de indivíduos



Fonte: O Autor (2018)

Pode-se identificar que as barras demonstram quantitativamente a presença da espécie selecionada. Existe também a possibilidade de classificar a quantidade dos animais

por classe de abundância, permitindo assim visualizar o gráfico de maneira mais sucinta, pois as barras possuem valores mais próximos. O Quadro 13 elenca as classes de abundância por quantidade.

Quadro 13: Classes de abundância

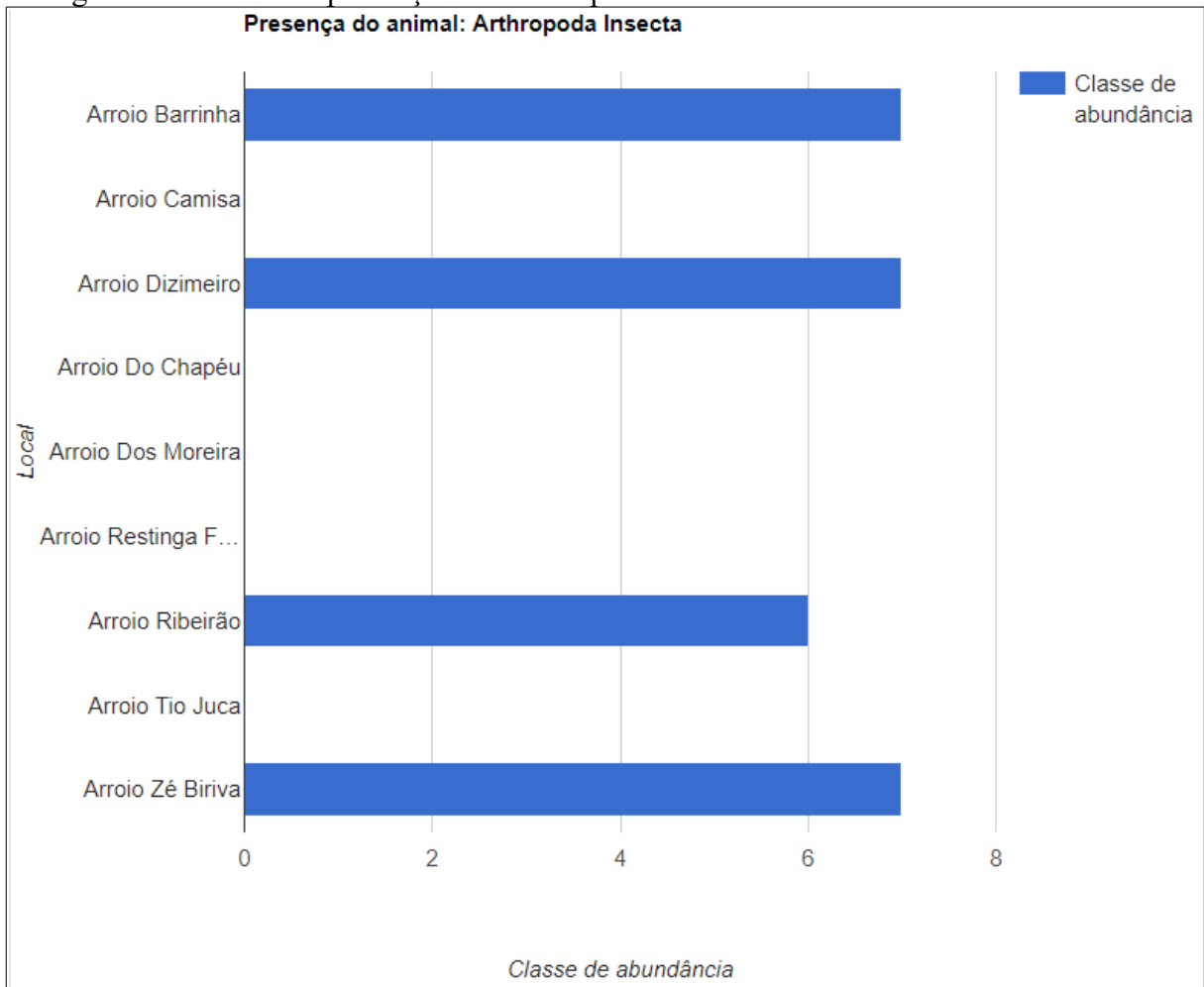
Número de indivíduos	Classe de abundância	Significado
1 e 2	1	Raro
3 a 10	2	Pouco
11 a 30	3	Pouco a médio
31 a 100	4	Médio
101 a 300	5	Médio a muito
301 a 1000	6	Muito
> 1000	7	Ocorrência massiva

Fonte: Adaptado de SCHÄFER(2013)

A Figura 44 representa uma execução para exemplificar a forma como é plotado um gráfico onde a quantidade dos indivíduos é representada por classes de abundância, através dos seguintes filtros de busca:

- **Data inicial:** 10/01/2001
- **Data final:** 31/12/2009
- **Agrupamento:** Espécie
- **Filo:** Arthropoda
- **Classe:** Insecta
- **Classificação:** Classe de abundância
- **Tipo de local:** Arroio
- **Demais filtros de busca não foram informados**

Figura 44: Gráfico da presença de animais por locais – Barras com classe de abundância



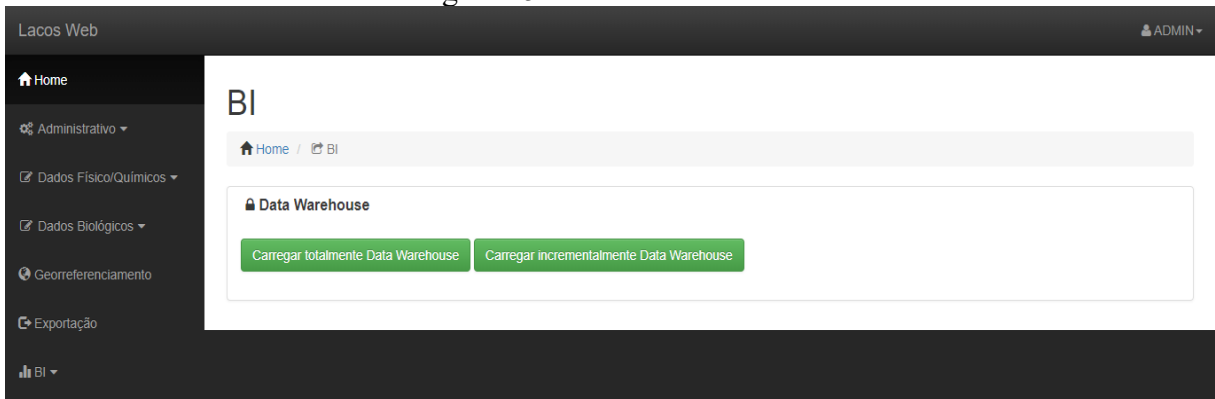
Fonte: O Autor (2018)

Nota-se que as barras apresentam valores de 0 a 7, os quais foram determinados conforme o Quadro 13, sendo a medida utilizada pelos pesquisadores do projeto LACOS.

### 5.1.3 BI – ADMINISTRATIVO

Esta parte do módulo é para uso restrito de usuários administradores e serve unicamente para realizar o ETL para o *DW*. A interface gráfica desta parte do sistema é composta por dois botões, os quais ao serem clicados chamarão a rotina que fará a extração, tratamento e carga total ou incremental dos dados no *Data Warehouse*, e alertará o usuário se a rotina obteve sucesso ou falha. A Figura 45 apresenta a interface gráfica desta parte da aplicação.

Figura 45: BI - Administrativo



Fonte: O Autor (2018)

## 5.2 QUESTIONÁRIO

Para medir a qualidade do software desenvolvido um questionário foi aplicado junto a pesquisadora Rosane Maria Lanzer, que é a responsável pelo projeto LACOS.

Referente a consulta de dados físico e químico:

1) Você compreendeu os filtros exigidos para realizar a consulta desejada? Teve alguma dificuldade?

Resposta: Não executei nenhuma consulta, somente visualizei o apresentado pelo acadêmico Lucas.

2) Após a consulta ser exibida em tela, você conseguiu rapidamente entender o(s) gráfico(s) exibidos? Se teve alguma dificuldade, cite quais.

Resposta: Os gráficos estão bem elaborados atendendo ao combinado quando de nossas conversas.

Referente a consulta dos dados de fauna:

3) Você compreendeu os filtros exigidos para realizar a consulta desejada? Teve alguma dificuldade?

Resposta: Não executei nenhuma consulta, somente visualizei o apresentado pelo acadêmico Lucas.

4) Após a consulta ser exibida em tela, você conseguiu rapidamente entender o(s) gráfico(s) exibidos? Se teve alguma dificuldade, cite quais.

Resposta: Da mesma forma que no caso anterior, os gráficos estão claros e atendem as necessidades estabelecidas no momento.

Avalie o módulo para BI nos seguintes quesitos.

5) Praticidade de uso:

[ ] ruim [ ] razoável [ ] bom [x] muito bom [ ] excelente [ ] não sei responder

6) Segurança nas informações exibidas:

ruim  razoável  bom  muito bom  excelente  não sei responder

7) Rapidez das consultas:

ruim  razoável  bom  muito bom  excelente  não sei responder

8) Agilidade de uso:

ruim  razoável  bom  muito bom  excelente  não sei responder

9) Você teria alguma sugestão de melhoria para o módulo apresentado?

Resposta: Tudo me parece muito bom, ressaltando que não atuei como usuário e sim como expectador

## 6 CONCLUSÃO

Para atender aos objetivos do presente trabalho, inicialmente teve que se tomar conhecimento e consciência de toda a importância que o projeto Lagoas Costeiras representa. Afinal é um projeto que estuda os recursos ecológicos e hídricos, provendo uma melhor identificação e gestão destes valiosos recursos naturais.

A partir da primeira versão desenvolvida do LACOS Web, em 2016, foi-se aos poucos incrementando as funcionalidades deste sistema, até todos os dados estarem organizados em uma base única e então que surgiu a necessidade de um módulo de BI a fim de dar aos pesquisadores *dashboards* de visualização capazes de mostrar dados estatísticos.

Com esta necessidade concretizada, iniciou-se a busca pelo conhecimento a cerca de *Business Intelligence*, abordando sua evolução, definição, conceito, aplicabilidade e arquitetura. Também foi de grande aporte elencar trabalhos relacionados, onde foi possível identificar elementos fundamentais que compõe o BI, tal como o *Data Warehouse* e uma boa interface com o usuário. No momento em que todo este material pesquisado foi compreendido, iniciou-se uma fase de levantamento de requisitos para o sistema, momento onde foi desenhado o modelo dimensional, diagramas de caso de uso, arquitetura de software e seus componentes.

Na fase de implementação, inicialmente foi instalado e configurado o ambiente de desenvolvimento conforme já havia sido implementado o restante do portal LACOS Web. O próximo passo então, foi a construção das tabelas do *Data Warehouse*, e o procedimento de ETL. A partir do momento em que o ETL foi finalizado, iniciou-se uma série de reuniões a fim de compreender e satisfazer as consultas a este DW. No mesmo ritmo em que ocorriam reuniões com a pesquisadora, foi feita a implementação dos *dashboards*, assim sendo o software teve três versões até ser apresentado e validado na última reunião realizada.

Os testes deste novo módulo foram limitados, pois como grande parte do acervo de dados do LACOS ainda não foi migrado, os cenários de teste ficaram restritos sobre o conjunto de dados que existem atualmente.

Portanto, agregado com o questionário aplicado a pesquisadora, conclui-se que o software conseguiu atender a necessidade dos pesquisadores, afinal, os *dashboards* conseguiram ser de fácil compreensão e apresentam informações necessárias aos usuários. Entretanto, cabe ainda ficar em observação, pois com a migração total do acervo de dados poderão surgir situações que não foram testadas.



## 6.1 TRABALHOS FUTUROS

Com o desenvolver do trabalho e das reuniões que foram necessárias até alinhar todos requisitos, surgiram uma série de sugestões de melhorias que podem vir a ser trabalhos futuros. Sugestões de gerar gráficos com as seguintes métricas podem vir a agregar bastante:

- IQ – Índice químico
- IQA – Índice de qualidade da água
- IET – Índice de estado trófico
- BMWP<sup>24</sup>
- Índices ASPT<sup>25</sup>

Outra sugestão de melhoria seria colocar mais opções de impressão do gráfico, deixando a geração do gráfico mais dinâmica e iterativa com o usuário. Tais opções como:

- Permitir modificar o tamanho dos gráficos
- Permitir *download* de cada gráfico gerado

Como todos os testes foram realizados com somente uma pequena parte dos dados que o LACOS possui, seria de grande valor a este módulo de BI, após todo o acervo de dados ser migrado para o LACOS Web, que fosse feito um novo questionário junto aos pesquisadores, a fim de detectar melhorias nos *dashboard* de visualização.

---

24 BMWP: Avaliação da Qualidade da Água Através dos Macroinvertebrados Bentônicos

25 ASPT: <https://aspt.net/>

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

- ALMEIDA, Ana Magela Rodriguez; CAMARGO, Sandra da Silva. Aplicando Técnicas de Business Intelligence sobre dados de desempenho Acadêmico: Um estudo de caso. In: Escola Regional de Banco de Dados. XI., 2015. Caxias do Sul. **Artigo**. Bagé, 2015. 98 – 107.
- AMARANTE, Raquel Raposo. **Business Intelligence no contexto de uma PME**. 2014. 56 f. Dissertação (Mestrado) - Curso de Economia, Universidade de Coimbra, Coimbra, 2014.
- ARONSON, Jay E. et al. **Business Intelligence – Um Enfoque Gerencial para a Inteligência do Negócio**. Porto Alegre: Bookman, 2009.
- CAYRES, Carlos Eduardo; OLIVEIRA, João Ricardo Dias de; MARINI, Alexandre. Business Intelligence na era da informação e as vantagens do Oracle na efetivação dessa tecnologia. Revista de Ciências Exatas e Tecnologia. São Paulo. Vol IV. Nº 4, 2009.
- COELHO, Felipe Zanellato; GOMES, Igor Meirelles; LANNES, Patricia de Aquino; SOUZA, Maria Alice V. Ferreira. de; BALDAM, Roquemar de Lima. BUSINESS INTELLIGENCE: SUPORTE À GESTÃO DO DESEMPENHO CIENTÍFICO. Encontro Nacional de Engenharia de Produção. XXXII., 2012. Bento Gonçalves, 2012. **Artigo**. Disponível em [http://www.abepro.org.br/biblioteca/enegep2012\\_TN\\_STO\\_164\\_958\\_20833.pdf](http://www.abepro.org.br/biblioteca/enegep2012_TN_STO_164_958_20833.pdf) > Acesso em: 15 nov. 2017.
- COSTA, Sérgio António Ramos da. **Sistema de Business Intelligence como suporte à Gestão Estratégica**. 2012. 135 f. Dissertação (Mestrado) - Curso de Mestrado em Engenharia e Gestão de Sistemas de Informação, Universidade do Minho, Braga, 2012.
- FRANCISCO, Vera Sofia Moreia. **Business Intelligence - Um Caso Prático**. 2009. 76 f. Dissertação (Mestrado) - Curso de Engenharia Informática e Computação, Faculdade de Engenharia da Universidade do Porto, Porto, 2009.
- FILHO, Wilson de Pádua Paula. **Engenharia de Software: Fundamentos, Métodos e Padrões**. 3ª ed. Rio de Janeiro: LTC, 2009
- FORTULAN, Marcos Roberto; FILHO, Eduardo Vila Gonçalves. **Uma proposta de aplicação de business intelligence no chão-de-fábrica**. Gestão e Produção, São Carlos, v. 12, n. 1, p.55-66, 25 fev. 2005.
- ELIAS, Diego. **A abordagem Top-Down e Bottom-Up no Data Warehouse**. Canal tech2014. Disponível em: <<https://canaltech.com.br/infra/a-abordagem-top-down-e-bottom-up-no-data-warehouse-21108/>>. Acesso em: 09 out. 2017.
- ELMASRI, Ramez; NAVATHE, Shamkant. **Sistema de banco de dados**. 6. ed. São Paulo: Pearson, 2010. 774 p.
- GOMES, António Boialvo Gomes. **Modelação de um Data Warehouse para a Direcção-Geral do Tesouro e Finanças e implementação de um Data Mart para o processo de**

**Gestão Patrimonial.** 2010. 99 f. Dissertação (Mestrado) - Curso de Mestrado em Estatística e Gestão de Sistemas de Informação, Universidade do Nova de Lisboa, Lisboa, 2010.

GOOGLE. **Google Charts..** Disponível em: <<https://developers.google.com/chart/interactive/docs/gallery>>. Acesso em: 28 mai. 2018.

GOUVEIA, António Roberto Taveira de Vasconcelos Pinto de. **Solução de Business Intelligence para Seguros.** 2013. 98 f. Dissertação (Mestrado) - Curso de Mestrado Integrado em Engenharia de Redes e Sistemas Informáticos, Universidade do Porto, Porto, 2013.

HERRERA, Jorge Anthony Félix. **Uso de Data Warehousing e Data Mining na busca de relações e conhecimento em um ambiente de comércio eletrônico.** 2003. 63 f. Dissertação (Mestrado) - Curso de Ciência da Computação e Matemática Computacional, Universidade de São Paulo, São Carlos, 2003.

INMON, William Harvey. **Como construir o data warehouse.** 2. ed. Rio de Janeiro: Campus, 1997.

LACOS. **Lagoas costeiras.** Disponível em: <<https://www.ucs.br/site/lacos3/>>. Acesso em 07 jun 2018.

LOH, Stanley. **BI na era do big data para cientistas de dados – indo além de cubos e dashboards na busca pelos porquês, explicações e padrões.** Porto Alegre, 2014.

KANASHIRO, Augusto. **Um data warehouse de publicações científicas: indexação automática da dimensão tópicos de pesquisa dos data marts.** 2007. 109 f. Dissertação (Mestrado) - Curso de Ciência da Computação e Matemática Computacional, Universidade de São Paulo, São Carlos, 2007.

KIMBALL, Ralph, and Monica Rosemberg. **Data warehouse toolkit.** São Paulo: Makron Books, 1998.

KIMBALL, Ralph; ROSS, Margy. **The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling.** 2. ed. Nova Iorque: John Wiley And Sons, Inc, 2002.

MARTINY, Ismael; RIBEIRO, Helena Graziottin. Módulo para Business Intelligence Compatível com Dispositivos Móveis para o Software de Gestão Empresarial Elementare. In: Escola Regional de Banco de Dados. XI., 2015. Caxias do Sul. **Artigo.** Caxias do Sul, 2015. 108 – 111.

PAULA, Danieli. **Business intelligence no auxílio da gestão da inovação: um estudo de caso utilizando sql server integration services e microstrategy.** Revista Interdisciplinar Científica Aplicada, Blumenau, v.10, n.2, p.69-92, TRII 2016. ISSN 1980-7031.

PRIMAK, Fábio Vinícius. **Decisões com B.I. (Business Intelligence).** Rio de Janeiro: Ciência Moderna, 2008.

RANJAN, Jayanthi. **Business intelligence: Concepts, components, techniques and benefits.** *Journal of Theoretical and Applied Information Technology.* V. 9, n.1, p. 60-70, 2009.

REZENDE, Solange Oliveira; PUGLIESI, Jaqueline Brigladori; MELANDA, Edson Augusto; PAULA, Marcos Ferreira de. Mineração de dados. In: REZENDE, Solange Oliveira, ed. **Sistemas Inteligentes – Fundamentos e Aplicações**, v. 1, (1 ed.), cap. 12, Barueri, SP: Manole, p.307-335, 2003.

SANTOS, Eliane Terezinha Lorenzi. **Estudo da implementação de OLAP para análise financeira**. 2006. 54 f. Monografia (Especialização) - Curso de Pós-graduação Especialização em MBA Gerenciamento em Banco de Dados II, Universidade do Extremo Sul Catarinense, Criciúma, 2006.

SANTOS, Maribel Yasmina, RAMOS, Isabel. **Business Intelligence – Tecnologias da Informação na Gestão de Conhecimento** (2 ed.). Lisboa, Portugal: FCA. 2009.

SCHÄFER, Alois Eduard; LANZER, Rosane Maria; SCUR, Luciana (Org.). **Atlas socioambiental dos municípios de Cidreira, Balneário Pinhal, Palmares do Sul**. Caxias do Sul, RS: EDUCS, 2013.

SILVA, Dhiogo Cardoso da. **Uma arquitetura de business intelligence para processamento analítico baseado em tecnologias semânticas e em linguagem natural**. 2011. 163 f. Dissertação – Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, 2011.

SILVA, Ernando. Uma abordagem utilizando business intelligence para apoiar o processo de tomada de decisão na gestão da evolução de serviços web. In: Simpósio Brasileiro de Bancos de Dados (SBBDD), 11. 2012, São Paulo. **Anais...** Porto Alegre, 2012. 69 – 74. Disponível em <<http://www.lbd.dcc.ufmg.br/bdbcomp/servlet/Evento?id=469>> acesso em: 15 nov 2017.

Silva, Ernando, Vollino, Bruno, Becker, Karin e Galante, Renata. (2012). **A Business Intelligence Approach to Support Decision Making in Service Evolution Management**. In: IEEE International Conference on Service Computing (SCC 2012)

TURBAN, Efraim, SHARDA, Ramesh, DELEN, Dursun. **Decision support and business intelligence systems**. 9. ed. Upper Saddle River, NJ, USA: Prentice Hall Press. 2010.

SOMMERVILE, Ian. **Engenharia de Software**; tradução Ivan Bosnic e Kalinka G. de O. Gonçalves; revisão técnica Kechi Hirama. —9. ed. —São Paulo: Pearson Prentice Hall, 2011.

Uno, Nairemilia Kuchauski Alves. **A importância do data warehouse no business intelligence**. ETIC – encontro de iniciação científica. Vol. 11. No 11. 2015. Disponível em: <<http://intertemas.toledoprudente.edu.br/revista/index.php/ETIC/article/view/4758>> Acessado em: 18/09/2017.

## ANEXO A – EXEMPLO DE IMPLEMENTAÇÃO DE GRÁFICOS UTILIZANDO GOOGLE CHARTS

Exemplo de geração de gráfico com Google Chart:

- **Referência para a biblioteca JavaScript:**

```
<script type="text/javascript" src="https://www.google.com/jsapi"></script>
```

- **Inclusão de dados para geração do gráfico:**

```
var data = new google.visualization.DataTable();
<?php
    if(isset($resultado)){
        foreach ($resultado as $row) {
            echo "data.addRow([[{$row->valor_avg}, -{$row-
>profund_local}]]);";
        }
    }
?>
var view = new google.visualization.DataView(data);
```

- **Código de Geração do Gráfico de Barras:**

```
var config = {
    'title': 'Título do gráfico',
    'width':500,
    'height':300,
    hAxis: { title: 'Título do eixo X'},
    vAxis: { title: 'Título do eixo Y'}
};
var chart = new
google.visualization.ColumnChart(document.getElementById('area_grafico'));
;
chart.draw(data, config); "
```