

**UNIVERSIDADE FEDERAL DO PAMPA**

**LEONARDO GAUER SCHULTE**

**APLICANDO TÉCNICAS DE BUSINESS INTELLIGENCE SOBRE  
DADOS DAS BIBLIOTECAS DA UNIPAMPA**

**Bagé  
2016**

**LEONARDO GAUER SCHULTE**

**APLICANDO TÉCNICAS DE BUSINESS INTELLIGENCE SOBRE  
DADOS DAS BIBLIOTECAS DA UNIPAMPA**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Engenharia de Computação como requisito parcial para a obtenção do grau de Bacharel em Engenharia de Computação.

Orientador: Sandro da Silva Camargo

**Bagé  
2016**

SC386 Schulte, Leonardo Gauer

Aplicando técnicas de Business Intelligence sobre dados das bibliotecas da UNIPAMPA / Leonardo Gauer Schulte. – dezembro, 2016.

90 f.: il.

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal do Pampa, Campus Bagé, Engenharia de Computação, 2016.

“Orientação: Sandro da Silva Camargo”.

1. Informação. 2. Conhecimento. 3. Suporte à decisão. 4. Gestão. 5. SpagoBI. I. Título.

**LEONARDO GAUER SCHULTE**

**APLICANDO TÉCNICAS DE BUSINESS INTELLIGENCE SOBRE  
DADOS DAS BIBLIOTECAS DA UNIPAMPA**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Engenharia de Computação como requisito parcial para a obtenção do grau de Bacharel em Engenharia de Computação.

Trabalho de Conclusão de Curso defendido e aprovado em: 03 de dezembro de 2016.

Banca examinadora:

---

Prof. Dr. Sandro da Silva Camargo  
Orientador  
UNIPAMPA

---

Prof. Dr. Leonardo Bidese de Pinho  
UNIPAMPA

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Ana Paula Lüdtke Ferreira  
UNIPAMPA

Dedico este trabalho às minhas três mães:  
Márcia Marli Gauer Schulte, Nelly Weigert  
Gauer e Mariza de Fátima Soares Flores e,  
ao meu pai: Ivan Köhler Schulte.

## AGRADECIMENTO

À Universidade Federal do Pampa, em especial, ao curso de Engenharia de Computação e todo seu corpo docente, que me proporcionaram as condições necessárias para que eu alcançasse meus objetivos.

Ao meu orientador Prof. Dr. Sandro da Silva Camargo, pela oportunidade e apoio na elaboração deste trabalho e também ao Prof. Dr. Leonardo Bidese de Pinho e à Prof<sup>a</sup>. Dr<sup>a</sup>. Ana Paula Lüdtke Ferreira, pelas contribuições.

À minha família, em especial aos meus pais Ivan Köhler Schulte e Márcia Marli Gauer Schulte, por me educarem e por me darem apoio, incentivo e todas as condições de realizar meus sonhos, minha irmã Luíse Gauer Schulte, pela amizade e por ser motivo de tanto orgulho, minha avó Nelly Weigert Gauer e meu tio Harto Cristiano Gauer, por me darem uma segunda casa onde tive educação, cuidado e muitos bons momentos e, minha namorada Julia Eisenhardt de Mello, pelo companheirismo em todos os momentos.

À minha outra família, Ildelfonso Flores, Mariza de Fátima Soares Flores, Eliezer Soares Flores e Filipe Soares Flores, por tamanha bondade e generosidade de fazer por mim o que ninguém mais faria: me acolher, cuidar e me fazer um dos seus quando eu era apenas um estranho, sem nunca esperar nada em troca.

Aos meus amigos, por tantos momentos de diversão e parceria, em especial, ao grande amigo César Huegel Richa, por estar sempre disposto a ajudar no que for preciso.

A todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigado.

## RESUMO

Atualmente, as organizações geram e armazenam grandes quantidades de dados referentes às atividades por elas realizadas. Ainda assim, têm pouca disponibilidade de informação e consequentemente, pouca disponibilidade de conhecimento. Isso acontece porque muitas vezes a quantidade de dados é tão grande que sua análise e processamento por técnicas convencionais é impossível. Nesse contexto, surge o *Business Intelligence* (BI): um conjunto de tecnologias de suporte à tomada de decisão que proporciona a gerentes, executivos e analistas decisões melhores e mais rápidas, com base na descoberta de conhecimento sobre grandes conjuntos de dados. Apesar de ter surgido no ambiente de negócios, o seu uso não se aplica apenas a esse contexto, podendo ser usado como apoio a decisões em diversos domínios. Instituições de ensino, como qualquer outro tipo de organização, têm por objetivo oferecer serviços de qualidade e bibliotecas acadêmicas tem papel fundamental nesse processo, pois buscam apoiar o ensino, extensão e pesquisa. Com isso, pôde-se identificar o potencial de aplicação do *Business Intelligence* nas bibliotecas da Universidade Federal do Pampa (UNIPAMPA). O presente trabalho apresenta as principais etapas da solução proposta, que visa otimizar a gestão das bibliotecas da UNIPAMPA através do desenvolvimento de um sistema de *Business Intelligence* capaz de fornecer conhecimentos relevantes a cerca das atividades desenvolvidas pelas bibliotecas, aos seus gestores. Esse trabalho iniciou com o levantamento de requisitos e pesquisa bibliográfica dos conceitos relativos ao assunto de interesse, de forma a embasar o projeto do *Data Warehouse* (DW) e *Data Marts* (DMs). O método adotado para a realização do *Extract, Transform, and Load* (ETL) e o uso das técnicas exploração dos dados. Durante o ETL, foi utilizada a ferramenta Talend Data Integration para extração dos dados das bases de dados da UNIPAMPA e carregamento dos mesmos para o esquema temporário do DW. Também foram criados gatilhos e funções com a linguagem PL/pgSQL responsáveis por transformar os dados e carregá-los para o esquema principal do DW. Com base nos *Key Performance Indicator* (KPIs), foram definidos os *datasets*, e com isso, foi possível a exploração e apresentação dos dados. Os resultados obtidos mostram que o BI é eficiente no suporte à extração de informações úteis, na forma de gráficos e relatórios, e descoberta de conhecimento capaz de apoiar o processo decisório e consequentemente, auxiliar a gestão da instituição.

**Palavras-chave:** Informação. Conhecimento. Suporte à decisão. Gestão. SpagoBI.

## ABSTRACT

Currently, organizations generate and store large amounts of data related to the activities they perform. However, they have small amounts of available information and therefore, small amounts of available knowledge. This occurs because often the large amount of data makes its analysis and processing impossible by conventional techniques. In this context, emerges Business Intelligence: a set of decision support technologies that provides to managers, executives and analysts a better and faster knowledge discovery in large data sets. Although it has emerged in the business environment, its use does not only apply to this context, but also to support decisions in several domains. Educational institutions, like any other type of organization, aim to provide quality services and academic libraries play a fundamental role in this process because aim support teaching, extension and research. Thus, it was possible to identify the application potential of Business Intelligence in the Federal University of Pampa libraries. The present work presents the main solution steps, which aim to optimize the management of UNIPAMPA libraries through the development of a business intelligence system capable of providing relevant knowledge to the activities developed for libraries and their managers. This work began with the requirements survey and bibliographical research of concepts related to interest of interest, so as to support the design of Data Warehouse (DW) and Data Marts (DMs). The method of extraction, transformation and loading (ETL) and the use of data mining techniques. During ETL, a Talend Data Integration tool was used to extract data from UNIPAMPA databases and upload them to the temporary DW scheme. Also, triggers and functions have been created with a PL/pgSQL language and are responsible for transforming the data and loading it into the main DW schema. Based on the Key Performance Indicators (KPIs), the data sets were given, and with this, it was possible to explore and present the data. The results show that BI is efficient without support to extract useful information, in the form of graphs and information, and the discovery of knowledge capable of supporting the decision process and, consequently, to help the institution management.

**Keywords:** Information. Knowledge. Decision support. Management. SpagoBI.

## LISTA DE FIGURAS

Figura 1 Hierarquia DIKW .....	19
Figura 2 Arquitetura básica de um sistema de BI .....	21
Figura 3 Componentes do processo de ETL .....	25
Figura 4 Modelos tipicamente usados em cada nível da arquitetura de BI.....	29
Figura 5 Representação de cubo OLAP .....	33
Figura 6 Modelo conceitual do banco de dados das bibliotecas da UNIPAMPA .....	38
Figura 7 <i>Data Mart</i> de empréstimos .....	40
Figura 8 <i>Data Mart</i> de exemplares .....	41
Figura 9 Modelo conceitual do <i>Data Warehouse</i> .....	41
Figura 10 Modelo lógico do esquema principal do DW .....	42
Figura 11 Modelo lógico do esquema intermediário do DW .....	43
Figura 12 Interface da Ferramenta Talend Data Integration .....	44
Figura 13 Interface do módulo QbE.....	46
Figura 14 Dashboard de acervo completo.....	47
Figura 15 Dashboard de acervo mostrando a quantidade de exemplares .....	48
Figura 16 Dashboard de acervo filtrado para o <i>campus</i> Bagé.....	48
Figura 17 Dashboard de acervo mostrando o custo dos exemplares.....	49
Figura 18 Empréstimos por <i>campus</i> no ano de 2015 .....	50
Figura 19 Empréstimos por horário no <i>campus</i> Bagé.....	51
Figura 20 Relatório de exemplares do <i>campus</i> Bagé nunca emprestados.....	51
Figura 21 Área de análise.....	52
Figura 22 Exemplares nunca emprestados por <i>campus</i> .....	53

## **LISTA DE TABELAS**

Tabela 1 Diferenças entre sistemas OLTP e OLAP .....	28
--	----

## LISTA DE SIGLAS

ACM	Association for Computing Machinery
BI	Business Intelligence
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DIKW	Data-Information-Knowledge-Wisdom
DM	Data Mart
DW	Data Warehouse
ER	Entidade - Relacionamento
ETL	Extract, Transform, and Load
GUI	Graphical User Interface
IBM	International Business Machines
IEEE	Institute of Electrical and Electronics Engineers
ISBN	International Standard Book Number
KPI	Key Performance Indicator
NTIC	Núcleo de Tecnologia da Informação e Comunicação
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
PDF	Portable Document Format
QbE	Query by Example
SAD	Sistemas de Apoio à Decisão
SADT	Structured Analysis and Design Technique
SGBD	Sistema Gerenciador de Banco de Dados
SQL	Structured Query Language
TAM	Technology Acceptance Model
TCC	Trabalho de Conclusão de Curso
UNIPAMPA	Universidade Federal do Pampa

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>13</b>
<b>1.1 Motivação</b> .....	<b>14</b>
<b>1.2 Objetivos</b> .....	<b>15</b>
<b>1.3 Metodologia</b> .....	<b>15</b>
<b>1.4 Estrutura do Trabalho</b> .....	<b>17</b>
<b>2 REVISÃO BIBLIOGRÁFICA</b> .....	<b>18</b>
<b>2.1 Definições e Conceitos Básicos</b> .....	<b>18</b>
2.1.1 Hierarquia <i>Data-Information-Knowledge-Wisdom</i> .....	18
2.1.2 Arquitetura .....	20
<b>2.2 Fonte de Dados</b> .....	<b>21</b>
2.2.1 Banco de Dados .....	21
<b>2.3 <i>Extract, Transform, and Load</i></b> .....	<b>23</b>
<b>2.4 <i>Data Warehouse</i></b> .....	<b>25</b>
2.4.1 Modelagem de Dados .....	29
<b>2.5 Recursos de Exploração</b> .....	<b>32</b>
2.5.1 <i>Online Analytical Processing</i> .....	32
2.5.2 <i>Data Mining</i> .....	34
<b>2.6 Componentes e Ferramentas</b> .....	<b>34</b>
<b>2.7 Desafios e Benefícios</b> .....	<b>35</b>
<b>2.8 Trabalhos Relacionados</b> .....	<b>36</b>
<b>3 SOLUÇÃO PROPOSTA</b> .....	<b>38</b>
<b>3.1 Obtenção dos dados</b> .....	<b>38</b>
<b>3.2 Identificação dos Indicadores de Desempenho</b> .....	<b>39</b>
<b>3.3 Projeto do <i>Data Warehouse</i></b> .....	<b>40</b>
<b>3.4 Processo de <i>Extract, Transform, and Load</i></b> .....	<b>43</b>
<b>3.5 Exploração dos Dados</b> .....	<b>46</b>
<b>3.6 Apresentação dos Conhecimentos Extraídos</b> .....	<b>47</b>
<b>4 CONSIDERAÇÕES FINAIS</b> .....	<b>54</b>
<b>REFERÊNCIAS</b> .....	<b>56</b>
<b>APÊNDICE A — DOCUMENTO DE REQUISITOS</b> .....	<b>59</b>
<b>APÊNDICE B — CARTA DE CIÊNCIA</b> .....	<b>67</b>

<b>APÊNDICE C — FUNÇÕES DE TRATAMENTO E CARGA DOS DADOS .....</b>	<b>69</b>
<b>APÊNDICE D — INTERFACES DA SOLUÇÃO .....</b>	<b>78</b>
<b>APÊNDICE E — DOCUMENTO DE AVALIAÇÃO DA SOLUÇÃO .....</b>	<b>88</b>

## 1 INTRODUÇÃO

O sucesso de uma organização depende das decisões que por ela são tomadas no seu dia a dia. Segundo Pereira e Fonseca (1997), tomar decisões é uma atividade que causa estresse aos gestores, pois sempre gera consequências e envolve riscos. Há fatores que influenciam diretamente o processo decisório, e parte deles existem antes mesmo de haver um problema de decisão a ser resolvido. Alguns fatores dizem respeito ao responsável pelo processo decisório, dentre os quais pode-se citar: cultura, classe social, sexo, religião, saúde, emocional, etc. Existem também fatores relativos ao meio na qual o gestor está inserido, alguns exemplos são: tecnologia disponível, normas da empresa, legislação, concorrência, notícias na mídia, tempo disponível, necessidade de produtos ou serviços de qualidade e com preços competitivos, etc. (BISPO, 1998).

Assim, é possível perceber que tomar decisões no contexto atual é uma atividade complexa. Para Cokins (2012), basear o processo decisório de uma instituição na intuição de seus gestores é perigoso, pois determinados executivos podem assumir riscos demasiados pensando que aquilo que funcionou no passado vai funcionar sempre, enquanto outros, podem ter dificuldade em assumir qualquer risco. Nenhum destes extremos é benéfico para a organização. Quando se assumem riscos, sendo estes baseados apenas na intuição dos tomadores de decisão, não existe qualquer garantia de que esta é uma boa escolha. Da mesma forma, uma empresa que não assume riscos dificilmente evolui, pois é importante que sejam feitos ajustes estratégicos para antecipar necessidades dos clientes, e também para reagir a táticas dos concorrentes.

Para que sejam feitas melhores escolhas, é importante baseá-las em informações relacionadas às atividades realizadas pela organização durante a sua história. Assim, é maior a probabilidade de esta contribuir positivamente para o crescimento da instituição (ALMEIDA et al., 2015). Para isso, é fundamental que os responsáveis pela tomada de decisões tenham acesso à consulta de informações de forma fácil, rápida e a qualquer hora, para que possuam o conhecimento necessário e sejam capazes de projetar o futuro de forma a visualizar tendências e diminuir riscos (TURBAN et al., 2009).

Com os atuais avanços da tecnologia, todas as organizações geram e armazenam grandes quantidades de dados referentes às atividades que realizam, mas muitas delas têm pouca disponibilidade de informação, logo, pouca disponibilidade de conhecimento. Isso ocorre porque muitas vezes a quantidade de dados é tão grande que torna impossível sua análise e processamento por técnicas convencionais e, assim se tornam pouco

aproveitadas pela organização que os detém. À medida que quantidade de dados armazenada se torna maior e cada vez mais dinâmica, a aplicação de técnicas de análise e processamento adequadas se torna cada vez mais importante (ANTONELLI, 2009).

Na década de 70, várias empresas e grupos de pesquisa começaram a desenvolver os Sistemas de Apoio à Decisão (SAD). De acordo com Sprague e Watson (1991), no princípio os SAD caracterizavam-se como sistemas computacionais interativos capazes de auxiliar o processo decisório de problemas não estruturados. Com o passar do tempo, essa definição foi ampliada, englobando qualquer sistema capaz de contribuir com o processo decisório, desde que possua recursos que facilitem o seu uso por pessoal não especializado em computação, sejam flexíveis a mudanças no ambiente da abordagem e combinem o uso de modelos ou técnicas analíticas a funções tradicionais de acesso e recuperação de informações.

Como resultado desse esforço, surge o *Business Intelligence* (BI) como um conjunto de técnicas que possibilitam a extração de informações relevantes desses conjuntos de dados, tornando-os úteis como suporte para tomada de decisões. Segundo Primak (2008), o termo BI surgiu na década de 80 no Gartner Group, empresa de consultoria que desenvolve tecnologias para auxiliar seus clientes a tomarem decisões melhores, e faz referência ao processo inteligente de coleta, organização, análise, compartilhamento e monitoramento de dados, gerando conhecimento para suporte a tomada de decisões no ambiente de negócios. Com o passar dos anos, essas técnicas não se restringem mais apenas ao ambiente de negócios e são usadas como apoio a decisões em ambientes dos mais diversos.

## **1.1 Motivação**

A necessidade de utilizar informações extraídas de dados para gerar conhecimento que auxilie os gestores no processo de tomada de decisões, aliado ao crescimento do emprego de técnicas de Business Intelligence fora do ambiente de negócios, mostrou o potencial da sua utilização em universidades. Instituições de ensino, em geral, têm por objetivo promover a formação humana e profissional de seus estudantes, preparando-os para uma atuação responsável e construtiva na sociedade, e para atingir esse objetivo é necessário que haja uma gestão qualificada.

Bibliotecas acadêmicas contribuem nesse processo, pois buscam apoiar o ensino e a aprendizagem. Para Zucca (2013), o cumprimento dessa missão está relacionado

ao consumo inteligente de recursos, qualidade do serviço, e o impacto da biblioteca nas pesquisas e na aprendizagem de estudantes.

Inserida neste meio, está a Universidade Federal do Pampa (UNIPAMPA), universidade federal brasileira de ensino superior criada em 2008, com dez *campi* distribuídos na metade sul e fronteira oeste do estado do Rio Grande do Sul, onde cada um dos *campi* possui uma biblioteca própria. Sendo assim, os recursos recebidos pela UNIPAMPA destinados às bibliotecas, devem ser divididos entre todas. Para isso, o ideal seria priorizar o repasse dos recursos aos *campi*, cursos e professores que melhor fazem uso dos mesmos.

O atual sistema de bibliotecas não é capaz de oferecer informações de forma eficiente, sobre quem são os *campi*, cursos e professores que melhor empregam os recursos recebidos. Os gestores das bibliotecas da UNIPAMPA possuem várias suspeitas sobre problemas de subutilização de recursos que ocorrem nas suas bibliotecas, contudo, não possuem informações para comprovar essas suspeitas e embasar decisões que possam erradicar ou, ao menos, minimizar esses problemas.

## 1.2 Objetivos

O objetivo deste trabalho é aplicar um conjunto de técnicas baseadas em *Business Intelligence* sobre dados das bibliotecas da Universidade Federal do Pampa, a fim de extrair conhecimento capaz de auxiliar no gerenciamento da instituição.

São objetivos específicos deste trabalho:

- Identificar o estado da arte em Sistemas de Apoio à Decisão e *Business Intelligence*;
- Selecionar uma ferramenta livre adequada para o problema;
- Projetar um Sistemas de Apoio à Decisão, baseado em ferramentas de *Business Intelligence* livres;
- Validar a solução.

## 1.3 Metodologia

O desenvolvimento deste trabalho foi dividido em duas etapas, sendo elas: planejamento e desenvolvimento da solução. O planejamento foi iniciado com a realização de uma pesquisa bibliográfica dos conceitos relativos ao *Business Intelligence*

e sua implementação, com o objetivo de obter referencial teórico para embasar o projeto da solução. Essa pesquisa foi feita por meio de livros, revistas e artigos científicos que abordam o assunto de interesse, buscados principalmente no *google scholar* e em bibliotecas digitais como a da *Institute of Electrical and Electronics Engineers (IEEE)* e *Association for Computing Machinery (ACM)* por exemplo, e também em material físico.

Nesta etapa, houve a necessidade de procurar os gestores da Universidade Federal do Pampa para apresentar os benefícios que o projeto poderia trazer para a instituição, e assim obter acesso aos dados relativos às bibliotecas da universidade e conseguir seu apoio na implementação da solução de *Business Intelligence*.

De posse dos dados, foi necessária a realização de uma reunião com a gestora responsável pelas bibliotecas da UNIPAMPA, com o objetivo de identificar suas necessidades informacionais. Nesta reunião, por meio de uma entrevista, foram definidos os requisitos do projeto, que são apresentados no apêndice A, e com base nela, posteriormente foram identificados os *Key Performance Indicators (KPIs)* mais importantes no processo de tomada de decisões.

A partir disso, iniciou-se o desenvolvimento da solução de *Business Intelligence*, com o projeto do *Data Warehouse*. Segundo Almeida et al. (2015), o DW é um depósito de dados com as informações consolidadas relativas ao banco de dados original. De acordo com Date (2004), ele é orientado a assunto, integrado, não volátil e variante no tempo, características essas que facilitam a análise de grandes volumes de dados. O DW possui um ou mais subconjuntos lógicos especializados em determinado assunto, estes são chamados *Data Marts*.

Para o projeto do *Data Warehouse*, foi adotada a metodologia *bottom-up* proposta por Kimball e Ross (2013), que sugere que sejam criados inicialmente os *Data Marts* para que posteriormente sejam unificados em um DW completo. O DW foi criado no Sistema de Gerenciamento de Banco de Dados (SGBD) PostgreSQL, por ser uma ferramenta *open source* e contar com muitos recursos desejáveis em um software desse propósito. Para facilitar seu povoamento, além do esquema principal que reúne os DMs multidimensionais, foi projetado no *Data Warehouse* um esquema intermediário modelado com base no Banco de dados da UNIPAMPA de forma a facilitar o processo de *Extract, Transform, and Load*.

Segundo Kimball et al. (2010), no processo de ETL os dados são retirados da suas fontes, depois integrados, pré-processados e preparados com o objetivo de melhorar sua qualidade. Isso foi feito em duas etapas: a primeira consiste no mapeamento dos

dados de interesse para o esquema intermediário do DW, isso foi feito com o auxílio da ferramenta Talend Data Integration, a segunda etapa consiste no povoamento do esquema principal do DW, etapa essa realizada por meio de funções e gatilhos desenvolvidos em *Structured Query Language* (SQL) no PostgreSQL. Nesta etapa os dados são organizados e modelados de acordo com as necessidades do projeto, além de serem eliminadas algumas inconsistências da base, tais como: completar dados, tratar valores nulos, e eliminar registros irrelevantes para análise.

Finalmente pôde-se explorar os dados armazenados no DW, analisando-os sob várias perspectivas diferentes usando técnicas de *Online Analytical Processing* (OLAP). Com isso, foi possível a realização de consultas para obtenção de gráficos e relatórios relativos ao funcionamento das bibliotecas. Esses gráficos e relatórios formam um *dashboard* desenvolvido no SpagoBI Server com a finalidade de apresentar visualmente os conhecimentos extraídos dos dados para os gestores da UNIPAMPA.

#### **1.4 Estrutura do Trabalho**

O presente trabalho está dividido em quatro capítulos. O capítulo 1 corresponde à introdução, motivação, objetivos e metodologia. No capítulo 2 é apresentado o referencial teórico relacionado ao *Business Intelligence*, necessário para embasar o projeto e a implantação da solução. Além disso são apresentadas as ferramentas utilizadas, os desafios e benefícios desse tipo de projeto e os trabalhos correlatos. No capítulo 3, é apresentada a solução proposta para as bibliotecas da UNIPAMPA, bem como a identificação dos indicadores de desempenho, projeto do *Data Warehouse*, processo de ETL, exploração e apresentação dos dados. No capítulo 4, são feitas as considerações finais, onde são analisados os resultados obtidos e apresentadas as expectativas para trabalhos futuros. Por fim, são apresentadas as referências do trabalho e os apêndices.

## 2 REVISÃO BIBLIOGRÁFICA

Neste capítulo serão apresentados, com base em uma revisão de literatura, alguns conceitos importantes relacionados ao processo de implementação de uma solução de *Business Intelligence* bem como definições de seus componentes de extração, tratamento, integração, armazenamento e exploração de dados.

### 2.1 Definições e Conceitos Básicos

Segundo Turban et al. (2009), *Business Intelligence* é um termo "guarda-cahuva" que inclui arquiteturas, ferramentas, banco de dados, aplicações e metodologias. Este termo é uma expressão livre de conteúdo, por isso pode ser compreendida de várias formas. Para Barbieri (2011), o conceito de BI pode ser entendido como a utilização de variadas fontes de informação para definir estratégias de competitividade nos negócios da empresa. Já Chaudhuri et al. (2011), dizem que o BI é um conjunto de tecnologias de suporte à tomada de decisão, que proporciona que gerentes, executivos e analistas tomem decisões melhores e mais rápidas com base na descoberta de conhecimento.

Os principais objetivos do BI são permitir acesso interativo aos dados, e facilitar sua manipulação, fornecendo aos gestores a capacidade de realizar análises adequadas. Portanto, sua definição está fortemente relacionada ao processo de transformação de dados em informação, e de informação em conhecimento, servindo como base para o processo decisório (TURBAN et al., 2009).

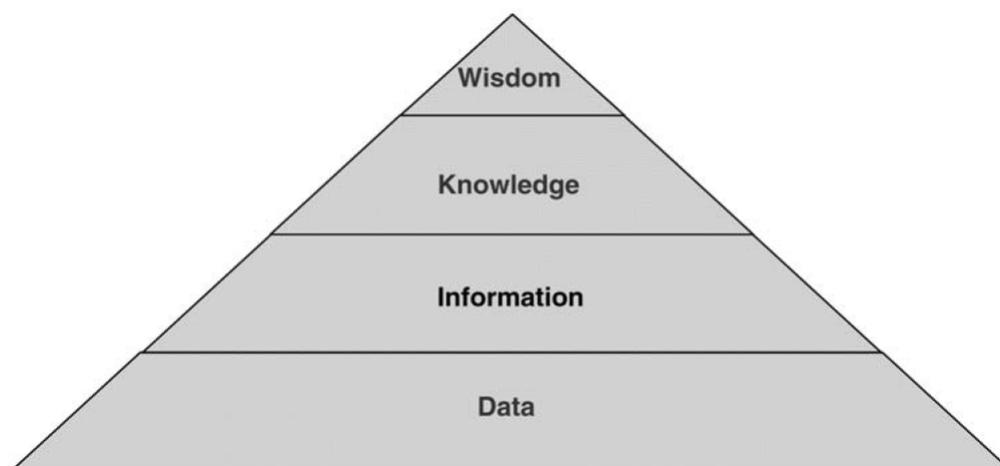
A importância do *Business Intelligence* no apoio ao processo decisório está fortemente relacionada a hierarquia *Data-Information-Knowledge-Wisdom* (DIKW), que visa justamente diferenciar dados, informação, conhecimento e sabedoria.

#### 2.1.1 Hierarquia *Data-Information-Knowledge-Wisdom*

Segundo Rowley (2007), a hierarquia DIKW, também conhecida como "pirâmide do conhecimento", foi proposta inicialmente por Ackoff (1989), que definiu as diferenças entre dados, informação, conhecimento, compreensão, inteligência e sabedoria, além de explorar processos associados à transformação entre estes elementos. Ackoff (1989), parte de uma descrição inicial onde cada elemento dessa hierarquia é um componente da

mente humana, e posteriormente apresenta uma perspectiva de sistemas de informação. Essa hierarquia foi refinada ao longo dos anos por diversos autores, sendo que atualmente ela costuma ser apresentada com os quatro níveis apresentados na figura 1 (ROWLEY, 2007).

Figura 1 – Hierarquia DIKW



Fonte: Rowley (2007)

Ackoff (1989 apud ROWLEY, 2007), define cada um dos níveis da hierarquia da seguinte forma:

- Dados são símbolos que representam as propriedades dos objetos, eventos e seu ambiente. Não tem qualquer significado ou valor porque são livres de contexto e interpretação. Eles são os produtos de observação, e podem existir em qualquer forma, utilizável ou não. É possível saber pouco a partir deles (ZELENY, 1987).
- Informação está contida em respostas a perguntas que começam com palavras como "quem", "o quê", "quando" e "quantos". Sistemas que geram, armazenam, recuperam e processam dados são capazes de produzir informação inferida a partir de dados. Informação pode ser útil, mas não necessariamente é. A diferença entre dados e informações é funcional e não estrutural.
- Conhecimento é a coleta adequada de informação útil. Torna possível a transformação da informação em instruções, e pode ser obtido mediante a transmissão de outro que o tem, por instrução, ou extraindo-o com a experiência.
- Sabedoria agrega o valor mental e requer julgamento. São aplicados valores éticos, o que a torna pessoal e faz com que seja inerente ao ator. A sabedoria é capaz de aumentar a eficácia das ações de quem a possui.

De posse desses conceitos, é possível observar que o *Business Intelligence* visa justamente partir do nível mais baixo da pirâmide rumo ao topo. Para isso, são utilizados alguns componentes que auxiliam nesse processo. Esses componentes formam a arquitetura do sistema de *Business Intelligence*, que é apresentada na seção 2.1.2.

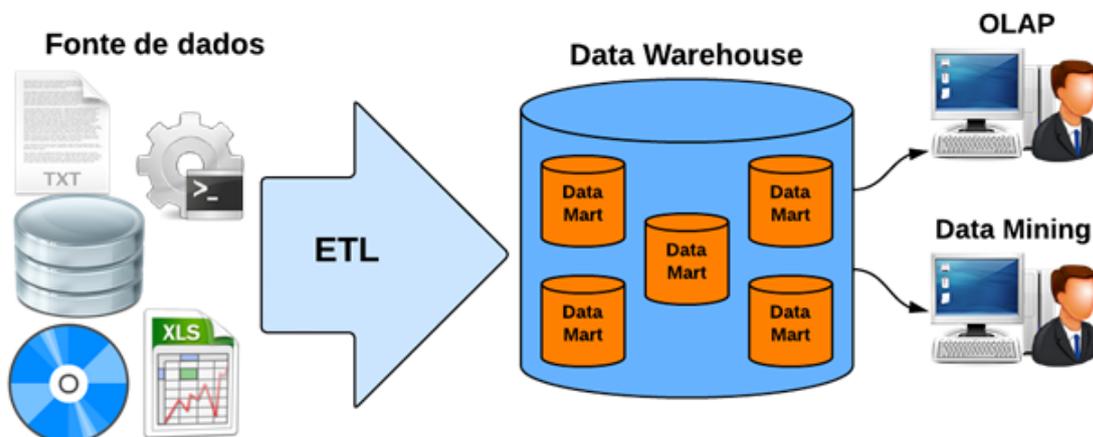
### 2.1.2 Arquitetura

Para Kimball e Ross (2013), a arquitetura básica que compõe um sistema de BI é formada pelos quatro componentes apresentados a seguir:

- Fontes de dados, que tratam dos sistemas de onde se originam os dados da empresa, podendo estar em formatos diversos (fontes heterogêneas), e também em lugares diferentes, havendo a necessidade de integração entre elas.
- *Extract, Transform, and Load*, que consiste na etapa de extração dos dados das fontes, sua integração, pré-processamento e preparação. Nesse processo, são eliminadas algumas inconsistências da base, tais como dados incompletos, valores nulos, e existência de registros irrelevantes para a análise.
- *Data Warehouse*, que é criado a partir do processo de ETL, armazenam as informações consolidadas relativas ao banco de dados original, já modeladas e preparadas para a exploração. O DW é composto por um ou mais subconjuntos lógicos, especializados em determinado assunto, de acordo com seu conteúdo. Esses subconjuntos são chamados de *Data Marts* (DMs).
- Área de apresentação, onde é feita a exploração dos dados. São usadas técnicas de análise de dados sobre as bases consolidadas para gerar informação de suporte a tomada de decisões, dentre as quais pode-se citar *Online Analytical Processing* e *Data Mining* como sendo as mais aplicadas atualmente.

A figura 2 ilustra a disposição dos componentes descritos acima.

Figura 2 – Arquitetura básica de um sistema de BI



Fonte: Canaltech (2016)

Essa arquitetura não obrigatoriamente precisa ser seguida. O analista de BI tem liberdade para projetar arquiteturas diferentes que melhor se adequem a contextos específicos. Porém, a arquitetura apresentada é muito citada na literatura pois é bastante genérica e se adapta muito bem à maioria das aplicações. As seções 2.2, 2.3, 2.4 e 2.5, apresentam o detalhamento dos componentes dessa arquitetura, e mostram como eles se relacionam.

## 2.2 Fonte de Dados

Fonte de dados refere-se ao local onde as empresas guardam seus registros operacionais. No passado esses registros eram armazenados em arquivos de papel, mas com o aumento da informatização das empresas, hoje a maioria delas guarda seus dados em formato digital, em planilhas, arquivos texto, e principalmente em bancos de dados.

### 2.2.1 Banco de Dados

Para Date (2004), um banco de dados é um sistema computadorizado de armazenamento de dados em forma de registros. Os usuários deste sistema podem executar operações de busca, inserção, modificação e remoção, sobre os registros.

Em meados da década de 60, todas as aplicações que precisavam guardar dados o faziam por meio de arquivos. Porém, à medida que a quantidade e a complexidade dos

dados aumentava, isso começou a se tornar problemático, pois era necessário um grande número de pessoas para desenvolver as aplicações e as manutenções e atualizações se tornavam extremamente custosas.

Por esse motivo, foram desenvolvidas pesquisas com o intuito de investigar novas maneiras de guardar dados, que apresentassem menores custos e maior eficiência. Como resultado desses esforços, na década de 70 o pesquisador da *International Business Machines* (IBM), Edgar Frank Codd, publicou um estudo sobre bancos de dados relacionais, onde ele imaginava que o usuário seria capaz de acessar dados por meio de comandos, o que anos mais tarde deu origem a *Structured Query Language* (GALASSI et al., 2013).

Bancos de dados se tornaram componentes essenciais no cotidiano da sociedade moderna. Na era da informática, da comunicação digital, em que dados podem ser compartilhados de um extremo a outro do planeta em segundos, a informação tornou-se o ativo mais importante das empresas. É comum encontrar-se bancos de dados com centenas de tabelas interrelacionadas com alguns milhões de registros e, por isso, o armazenamento precisa ser feito de forma eficiente (DATE, 2004). Segundo Elmasri (2003), os bancos de dados representam um papel crítico em quase todas as áreas em que os computadores são utilizados, incluindo negócios, comércio eletrônico, engenharia, medicina, direito, educação e as ciências da informação.

Os Sistemas Gerenciadores de Banco de Dados são coleções de *softwares* que permitem aos usuários criar e manterem um banco de dados, facilitando o processo de definição, construção, manipulação e compartilhamento de dados entre vários usuários e aplicações. A construção é o processo de armazenar os dados em alguma mídia apropriada controlada pelo SGBD. A manipulação é feita pelas funções *select*, *insert*, *update* e *delete*, capazes de realizar respectivamente, consultas, inserções, atualizações e exclusões de registros no banco. O compartilhamento permite aos múltiplos usuários e programas acessarem de forma concorrente o banco de dados (ELMASRI, 2003).

Por oferecerem uma interface de manipulação e compartilhamento de dados os SGBDs trazem consigo diversos outros benefícios. Para Date (2004), as principais vantagens no uso de bancos de dados são:

- Redundância pode ser reduzida: Em sistemas sem bancos de dados cada aplicação precisa ter os seus próprios arquivos. Esse fato pode levar a uma considerável redundância de dados, e representar considerável desperdício de armazenamento.

- Inconsistências podem ser evitadas: Se houver redundância não controlada, duas aplicações que utilizam um mesmo arquivo, em determinado momento podem estar utilizando versões diferentes do mesmo, uma vez que o SGBD não tem conhecimento sobre essa redundância.
- Suporte a transações: Uma transação é uma unidade lógica de trabalho que envolve operações de atualização em bancos de dados. Um exemplo típico envolve a transferência de um determinado valor de uma conta bancária A para uma conta bancária B. É claro que, para que essa transação ocorra corretamente, duas atualizações são necessárias. Assim, o suporte a transações garante que ambas sejam realizadas, ou então nenhuma delas, ainda que o sistema venha a falhar (por uma falta de energia por exemplo) em meio ao processo.
- Integridade: Mesmo quando não há redundâncias, ainda podem haver informações incorretas. Por um erro comum de digitação, o banco poderia mostrar que um empregado trabalhou 400h em uma semana, o que claramente é impossível. Para esse tipo de problemas é possível implementar restrições de integridade a serem executadas sempre que ocorrer uma operação de atualização.
- Segurança: Sob a orientação apropriada do administrador, o SGBD pode assegurar que o acesso ao banco se dê através de canais apropriados, e pode definir restrições de segurança a serem verificadas sempre que houver tentativa de acesso a determinados dados. Podem ser definidas inclusive restrições diferentes, dependendo do tipo de acesso e do usuário que está tentando acessar.

Apesar dos benefícios apresentados, muitas vezes esses recursos não são explorados de forma correta pelos projetistas e administradores de bancos de dados, causando danos à qualidade dos dados que armazenam. Por isso, se faz necessário o processo de ETL, para preparar os dados e possibilitar a extração de conhecimento dos mesmos.

### ***2.3 Extract, Transform, and Load***

O processo de *Extract, Transform, and Load* visa a sistematização do tratamento e limpeza dos dados oriundos das diversas fontes de dados para a inserção em um DW. Segundo Turban et al. (2009), o ETL é a principal etapa do processo de concepção de *Data Warehouses*, e componente central de qualquer projeto centrado em dados, podendo

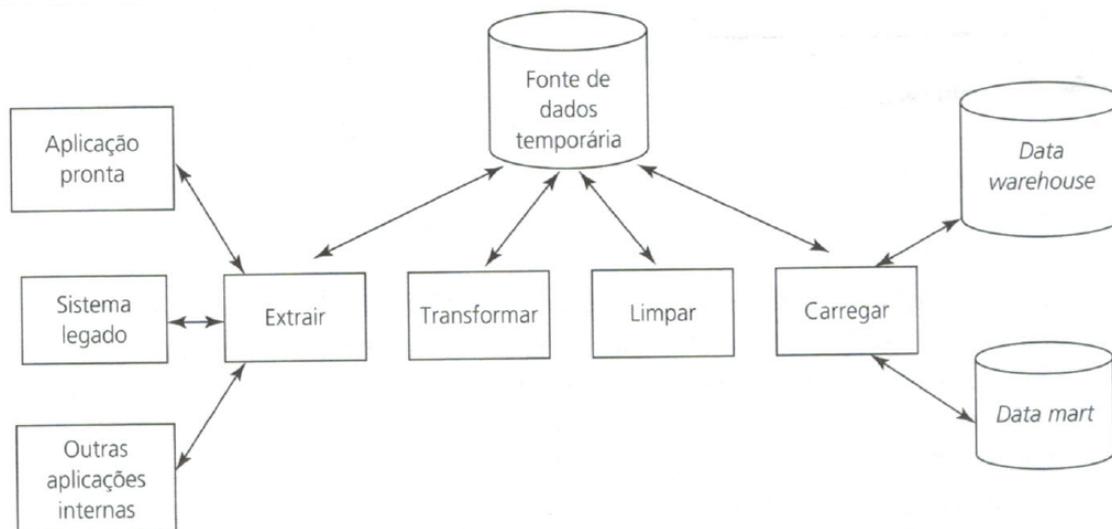
consumir até 70% do tempo do mesmo.

Na etapa de extração dos dados se faz necessária a definição dos KPIs. Segundo Parmenter (2015), esses indicadores são utilizados principalmente para mensurar aspectos quantitativos de fenômenos determinantes para o sucesso da organização. Os KPIs são ferramentas de gestão, que facilitam a transmissão da missão de uma empresa aos funcionários que não estão em cargos de gestão, a partir da definição de metas que devem ser alcançadas em determinado período. Assim, a definição dos KPIs mostra quais os dados devem ser explorados e monitorados pela solução de BI e, possibilita que o processo de ETL seja mais rápido e resulte em um DW mais compacto, uma vez que dados irrelevantes para análise podem ser ignorados antes mesmo de passar pela etapa de transformação.

De acordo com Turban et al. (2009), os dados extraídos de suas fontes originais são carregados em uma base de dados temporária. Depois disso, é feita a transformação, que consiste na correção de inconsistências e conversão dos dados, para o formato compatível com o *Data Warehouse*. Existem ferramentas de transformação que oferecem uma *Graphical User Interface* (GUI) que auxilia no desenvolvimento e manutenção de *Data Warehouses*, e facilitam o processo de integração e carga de dados, mas esse processo também pode ser realizado por meio do desenvolvimento de *scripts* usando linguagens como Python, PL/SQL, C, etc.

A figura 3, ilustra os componentes do processo de ETL. Nela, os dados são extraídos das diversas fontes e carregados na área de preparação, depois processados e carregados no *Data Warehouse*.

Figura 3 – Componentes do processo de ETL



Fonte: Turban et al. (2009)

## 2.4 Data Warehouse

Segundo Turban et al. (2009), um *Data Warehouse* que é repositório com dados de interesse especial preparado para dar suporte a aplicações de tomada de decisão. Para Inmon (2005), Sezões et al. (2006) e Turban et al. (2009), um DW possui as seguintes características:

- Organizado e orientado por assunto: os dados são organizados por temas chave da organização. Normalmente são apresentados de forma compartimentada, de acordo com as necessidades específicas dos usuários finais do sistema. Visam a análise de dados específicos que sejam relevantes no processo decisório.
- Integrado: deve ser uma fonte de dados única que com dados referentes a todo o processo de negócio. Os dados referentes às diversas fontes operacionais da empresa são unidos propositalmente no processo de ETL, proporcionando uma visão única e abrangente, de forma a responder a qualquer questão imposta pelos gestores.
- Catalogado temporalmente: o *Data Warehouse* tem como principal objetivo fornecer informações atuais que sejam válidas sob uma perspectiva histórica, e por isso deve possuir uma dimensão temporal capaz de situar o objeto de estudo.

- Não volátil: deve apresentar dados estáveis, ou seja, manter dados históricos sem a possibilidade de exclusões. Conseqüentemente, no ambiente de *Data Warehouse* podem haver três tipos de operações: o carregamento inicial dos dados, atualizações (inserção de novos registros), e o acesso aos dados.
- Relacional/Dimensional: pode ser modelado de forma relacional ou dimensional (ou multidimensional) de acordo com o tipo de aplicação.
- Baseado na *Web*: *Data Warehouses* são criados para apresentar um ambiente informatizado eficiente em aplicações *Web*. Possuem arquitetura cliente/servidor proporcionando acesso fácil aos usuários.
- Inclui metadados: um *Data Warehouse* contém metadados (dados sobre dados) sobre como os dados estão organizados e como usá-los de forma eficiente.

O *Data Warehouse* possui subdivisões lógicas internas chamadas *Data Marts*. Os *Data Marts* armazenam subconjuntos dos dados do *Data Warehouse*. Assim cada setor da organização pode ter um ou mais *Data Marts* que contenham os assuntos de interesse, e a exploração pode ser feita de forma direcionada. Kimball e Ross (2013) define *Data Marts* como repositórios de dados multidimensionais, mais específicos do que os *Data Warehouses*, que reúnem um conjunto de tabelas dimensionais de suporte a um determinado processo de negócio.

Existem na literatura duas principais abordagens para implementação de *Data Warehouses*. A primeira, sugerida por Inmon (2005), baseia-se numa metodologia *top-down*, que consiste na definição do esquema global do *Data Warehouse*, para que depois sejam implementados os *Data Marts* de acordo com as necessidades informacionais, e características de negócio da organização. A segunda abordagem de implementação é uma metodologia *bottom-up*, sugerida por Kimball e Ross (2013). Esta, baseia-se na experimentação e desenvolvimento de protótipos. Nela, são desenvolvidos inicialmente esquemas individuais de cada *Data Mart*, com base nas necessidades do negócio. Estes *Data Marts* devem ser modelados tendo em vista a futura unificação dos mesmos, de modo a formar um *Data Warehouse* completo.

A abordagem de Inmon (2005), implica mais tempo de desenvolvimento e maior risco de não ser concluída dentro do prazo, pois exige que o projetista tenha uma visão completa do negócio como um todo para a criação do *Data Warehouse*, enquanto a abordagem de Kimball e Ross (2013), por ser incremental, é mais simples e rápida, e geralmente fornece resultados mais tangíveis, mas carece de uma visão conjunta de todo o sistema a ser desenvolvido (VERCELLIS, 2009 apud COSTA, 2012).

Segundo Chaudhuri e Dayal (1997 apud COSTA, 2012), Os *Data Warehouses* ficam separados das bases de dados operacionais das organizações, pois possuem finalidades diferentes. As bases de dados operacionais, conhecidas como sistemas *Online Transaction Processing* (OLTP), concebidos para registrar todas as operações do cotidiano de uma organização, por meio das operações de inserção, modificação e eliminação de registros na base de dados em um determinado período de tempo.

Já os sistemas de *Data Warehouse* são projetados para dar suporte às decisões das organizações, e portanto são considerados sistemas analíticos, conhecidos pelo uso de *Online Analytical Processing* (OLAP), uma vez que apresentam informação consolidada, histórica e resumida, sendo este um fator importante. As características do *Data Warehouse* estão relacionadas ao fato de este integrar informação referente a um determinados assuntos da organização, caracterizando-a como um todo. A tabela 1 sintetiza as principais diferenças entre os dois tipos de sistemas.

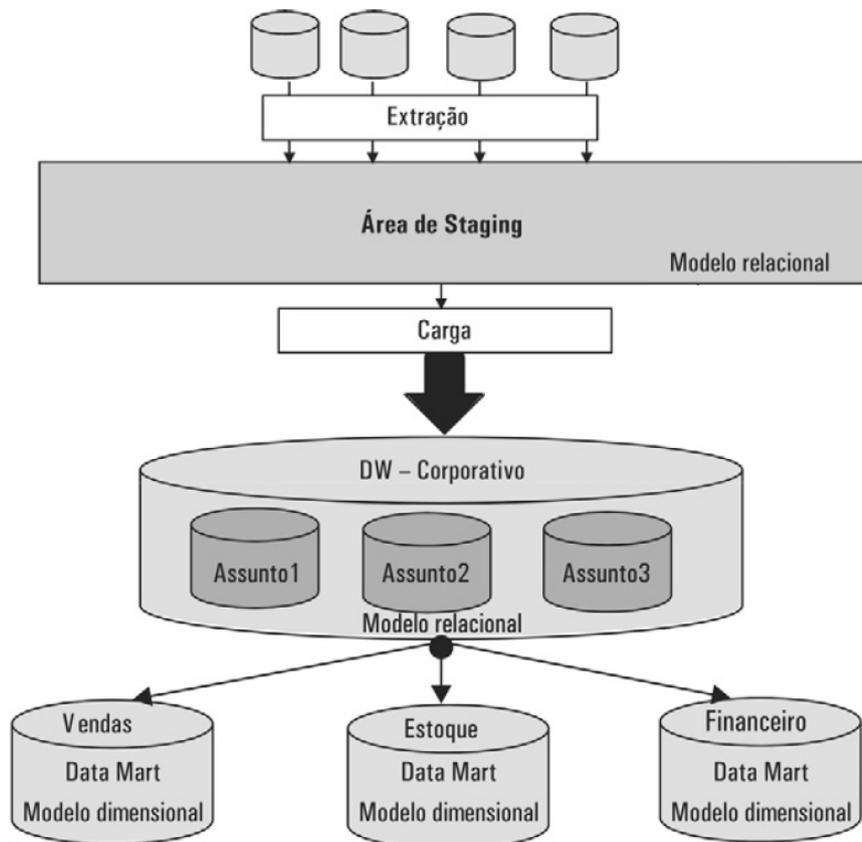
Tabela 1 – Diferenças entre sistemas OLTP e OLAP

Características	Sistemas OLTP	Sistemas OLAP
<b>Fonte de Dados</b>	- OLTP são as fontes originais dos dados;	- Consolidação de dados; - Dados OLAP surgem de vários sistemas OLTP;
<b>Propósito dos Dados</b>	- Executar e controlar tarefas fundamentais do negócio;	- Ajudar no planeamento, resolução de problemas e suporte à decisão;
<b>Tipo de Dados</b>	- Revelam o momento dos acontecimentos nos processos de negócio; - São atómicos, normalizados, atualizados e isolados;	- Evidenciam várias perspetivas (multidimensionais) das atividades organizacionais; - São históricos, sumarizados, multidimensionais e integrados;
<b>Inserções e Atualizações</b>	- Curtas e rápidas; - Atualizações efetuadas pelos utilizadores;	- Periódicas e de longa duração;
<b>Acessos/ Consultas</b>	- Leitura/escrita; - Consultas/transações padronizadas e simples (envolvem poucas tabelas e retornam poucos registos);	- Leitura na maior parte das vezes; Consultas complexas que envolvem agregações (envolvem várias tabelas e retornam muitos registos);
<b>Velocidade de Processamento</b>	- Normalmente muito rápido para inserir, apagar e alterar, menos rápido para consultar;	- Depende da quantidade de dados envolvidos; - Está otimizado para processamento de questões;
<b>Requisitos de Espaço</b>	- Pode ser relativamente pequena caso os dados históricos não sejam guardados;	- Requer mais espaço devido à existência de agregação dos dados e história. - Requer mais índices do que os Sistemas OLTP;
<b>Conceção da Base de Dados</b>	- Altamente normalizados e com muitas tabelas; - Orientado às aplicações; - Otimizados para atualizações;	- Normalmente desnormalizados e com menos tabelas; - Orientados aos assuntos; - Otimizados para o processamento de questões;
<b>Utilização</b>	- Repetitiva (diariamente).	- <i>Ad-hoc</i> .

Fonte: Costa (2012)

Outro fator que diferencia os bancos de dados OLTP, *Data Warehouses* e *Data Marts*, é o tipo de modelagem de dados comumente adotados em cada um. A figura 4, a seguir mostra uma alternativa de arquitetura de BI, onde é possível observar o tipo de modelagem adotada em cada nível.

Figura 4 – Modelos tipicamente usados em cada nível da arquitetura de BI



Fonte: Adaptado de Barbieri (2011)

O *Data Warehouse* tem fundamental importância para o sucesso de aplicações de *Business Intelligence*, uma vez que oferece dados consolidados e estruturados para tirar o melhor proveito dos recursos de exploração de dados. Grande parte desses benefícios provêm do tipo de modelagem que é utilizada no seu projeto. A seção 2.4.1 apresenta mais detalhes referentes à modelagem de dados, bem como os principais aspectos das abordagens relacionais e dimensionais e diferenças entre eles.

### 2.4.1 Modelagem de Dados

Para Heuser (2009), um modelo tem por objetivo representar de forma abstrata o mundo real, sendo que para um modelo de dados são levados em conta apenas os objetos sobre os quais deseja-se manter informações. Um modelo de dados corresponde à descrição formal da estrutura de um banco de dados. Nele são definidas quais as informações a serem guardadas e seus respectivos tipos, bem como as relações que

existem entre eles.

Para o mesmo autor, bancos de dados podem ser descritos em níveis de abstração distintos. Modelos de propósito geral, que visam apresentar apenas a organização do banco, dispensam maiores detalhes sobre a representação em meio físico das informações, e por isso apresentam um nível alto de abstração. Porém, modelos utilizados para atividades mais específicas, como por exemplo, otimizar a performance de acesso, demandam mais detalhes da organização interna do banco, e portanto são menos abstratos.

No projeto de banco de dados, normalmente são considerados dois níveis de abstração, o do modelo conceitual e o do modelo físico. Para Teorey et al. (2005), o modelo conceitual é o principal componente do projeto lógico do banco de dados. Para esses modelos, utiliza-se diagramas de esquemas para descrever a estrutura do banco de dados independentemente do SGBD na qual será instalado. São comprovadamente os mais populares e usados hoje em dia devido ao seu alto grau de abstração, proporcionando grande simplicidade e legibilidade. Os modelos físicos, são baseados nos detalhes do Sistema Gerenciador de Banco de Dados escolhidos para implementar o banco de dados, e por isso, geralmente são usados apenas por profissionais que visam otimizar a performance do banco.

Segundo Jardim et al. (2015), existem duas principais abordagens para a modelagem conceitual, a modelagem relacional e a dimensional, que serão descritas a seguir.

### *Modelo Relacional*

Segundo Barbieri (2011), o modelo relacional desenvolvido pelo britânico Edgar Frank Codd, tratava os dados de maneira puramente matemática. Nesse modelo, os registros eram tratados como entes de conjuntos e sujeitos a manipulações da álgebra relacional. Esse modelo desenvolvido por Codd foi aprimorado em 1976, por Peter Chen, que apresentou um modelo baseado em entidades e relacionamentos. O modelo Entidade - Relacionamento (ER) é descrito em formato de diagrama, emprega retângulos para representar entidades, e usa objetos em forma de losango para representar os vários tipos de relacionamentos, que são diferenciados por números ou letras colocadas nas linhas que conectam os losangos aos retângulos (HEUSER, 2009).

Para Heuser (2009), uma entidade representa um conjunto de objetos da realidade modelada, sobre os quais deseja-se guardar informações. Teorey et al. (2005), dizem que elas normalmente representam uma pessoa, lugar, coisa ou evento de interesse

informativo. Para o mesmo autor, relacionamentos não possuem existência física ou conceitual, eles representam associações do mundo real entre uma ou mais entidades. Esse termo é comumente utilizado para indicar grau de conectividade entre ocorrências de entidade, podendo ser: um-para-um, um-para-muitos e muitos-para-muitos.

Os modelos relacionais geralmente possuem muitas entidades e relacionamentos entre elas, pois cada objeto da realidade é representado por uma entidade diferente. Assim, essa metodologia proporciona uma aproximação muito grande da realidade modelada, e por isso, é a mais usada atualmente em sistemas OLTP.

### *Modelo Dimensional*

A modelagem dimensional, também chamada de multidimensional por vários autores, é definida por Costa (2012), como um modelo de dados que proporciona uma melhor otimização do sistema, no que diz respeito ao processamento de consultas complexas. Diferente dos bancos de dados relacionais que buscam a dispersão das informações para uma melhor representação da realidade, os bancos de dados dimensionais, tendem a realizar grande agregação de informações de forma a simplificar as consultas a grandes quantidades de dados. Para Turban et al. (2009), em um banco de dados multidimensional os dados estão organizados especialmente para permitir uma análise sob diferentes perspectivas de forma rápida e fácil. Criar e manter esse tipo de repositório de dados agrega custos que devem ser comparados com os benefícios agregados.

Segundo Barbieri (2011), dois tipos de tabelas compõem esse modelo:

- **Tabela Fato:** Possui característica quantitativa dentro do DW. A partir dela são extraídas as métricas que são cruzadas com os dados das Dimensões. Armazena medições necessárias para avaliar o assunto pretendido, sendo que cada fato pode armazenar uma ou mais medidas, que constituem os valores objetos da análise. Normalmente armazenam muito mais registros do que as tabelas dimensão e, contêm dados aditivos (manipulados por soma, média, contagem, etc.) e relativamente estáticos. Originam-se das entidades encontradas no modelo relacional que representam ações, eventos, acontecimentos, enfim, fatos que desejamos registrar.

- Tabelas dimensão: Possui característica descritiva dentro do DW. Ela qualifica as informações provenientes da tabela Fato. Através dela é possível analisar os dados sob múltiplas perspectivas. As tabelas dimensão têm uma relação um para muitos com a tabela fato e geralmente tem a função de situar os registros correspondentes a ele na tabela fato, respondendo perguntas como: "quem?", "quando?", "onde?", "o que?", etc. Possuem múltiplas colunas de informação, e devem ser entendidas como as tabelas que realizam os filtros de valores aplicados na manipulação dos fatos. Normalmente estão niveladas em hierarquia, e são agrupadas nos resultados das consultas.

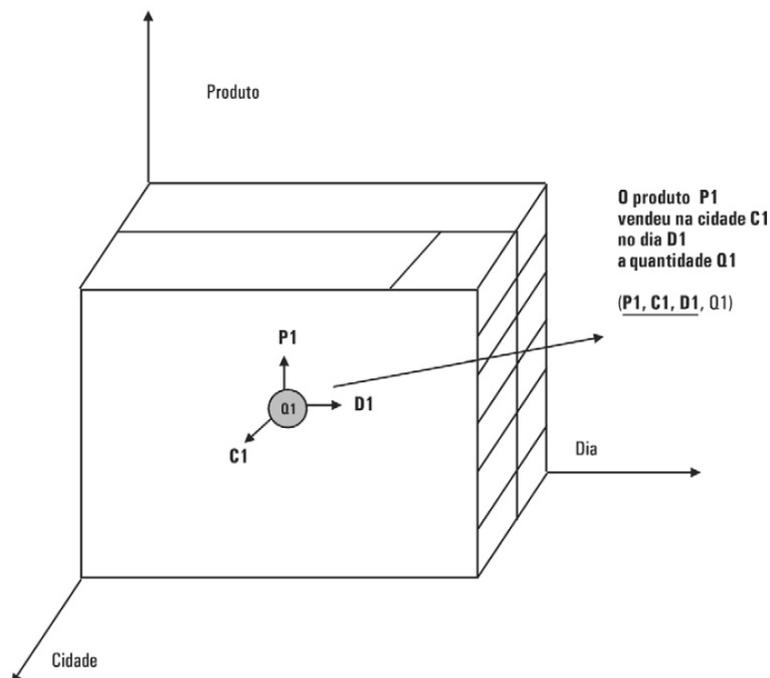
## **2.5 Recursos de Exploração**

Após concluído o processo de concepção do *Data Warehouse* e dos *Data Marts*, pode-se dar início à exploração dos dados. Para essa etapa, dois tipos de recursos destacam-se como os mais utilizados: *data mining* e OLAP. Ambos descritos nas seções 2.5.1 e 2.5.2.

### **2.5.1 Online Analytical Processing**

Um sistema de processamento analítico (OLAP), é uma tecnologia de exploração de dados que opera sobre repositórios de dados multidimensionais, ou seja, *Data Warehouses* e *Data Marts*, possibilitando uma visualização dos dados sob várias perspectivas, de maneira a apoiar as organizações na descoberta de conhecimento a partir da análise de dados (COSTA, 2012). Na literatura, se encontra muito o termo *cubo OLAP*, isso, porque a visão multidimensional proporcionada pelo OLAP, pode ser representada como uma estrutura em formato de cubo, como pode ser observado na figura 5. Nela está representado um produto P1, na cidade C1 e no dia D1, onde P1, C1 e D1 são as dimensões de um fato venda de quantidade Q1 (BARBIERI, 2011).

Figura 5 – Representação de cubo OLAP



Fonte: Barbieri (2011)

Para Turban et al. (2009), o termo OLAP refere-se a uma variedade de atividades executadas por usuários finais dos sistemas. Enquanto um *Data Warehouse* é utilizado para armazenar a informação, os sistemas OLAP oferecem recursos de modelagem, análise e visualização de grandes conjuntos de dados, proporcionando geração de respostas a consultas, solicitação e execução de relatórios e gráficos, realização de análises estatísticas e construção de apresentações visuais.

Segundo Turban et al. (2009) e Barbieri (2011), existem várias arquiteturas de OLAP, dentre as quais destacam-se o MOLAP, ROLAP e HOLAP. A seguir, são apresentadas as principais características de cada uma.

- OLAP multidimensional (MOLAP): Quando o OLAP é implementado por meio de bancos de dados multidimensionais especializados, os dados são organizados em forma de cubos que o usuário pode rotacionar. Com MOLAP as consultas são mais rápidas pois a consolidação já foi realizada previamente.
- OLAP relacional (ROLAP): Através de consultas SQL, ele extrai os dados de bancos de dados relacionais e cria dinamicamente visões multidimensionais sobre os dados. Tende a ser usado em estruturas que contém grande número de atributos que não podem ser colocados facilmente em uma estrutura de cubos.

- *Hybrid OLAP (HOLAP)*: Combina e integra as duas arquiteturas citadas anteriormente, beneficiando-se dos seus pontos fortes e minimizando as suas fraquezas, nomeadamente a escalabilidade da arquitetura ROLAP e da capacidade de maior processamento e velocidade da arquitetura MOLAP. Faz uso de estruturas relacionais para os dados de maior granularidade e as estruturas dimensionais nativas são dedicadas ao armazenamento de agregados.

### 2.5.2 Data Mining

Para Turban et al. (2009), *data mining* é uma classe de análise de informações que aplica tecnologias de reconhecimento de padrões e métodos estatísticos na busca de relações e padrões ocultos em uma coleção de dados. Segundo Barbieri (2011), o processo de data mining possui diferenças com relação ao OLAP. Enquanto o OLAP opera com fatos em dimensões variadas, para responder a perguntas predefinidas, as técnicas de *mining* busca algo além da interpretação dos dados existentes, ela visa, fundamentalmente, inferir possíveis comportamentos sobre os fatos de interesse, através da busca por correlações não explicitadas em meio a grande quantidade de dados de um *Data Warehouse*.

De acordo com Barbieri (2011), os recursos apresentados nesta seção, são os mais usados atualmente para exploração de dados em sistemas de *Business Intelligence*. Eles oferecem vários recursos, e são responsáveis pela geração de informações capazes de apoiar os gestores das organização no processo de tomada de decisões.

## 2.6 Componentes e Ferramentas

Existem atualmente no mercado diversas ferramentas que implementam os conceitos de Banco de dados e *Business Intelligence* apresentados nas seções anteriores. Grandes empresas do ramo de tecnologia como a Microsoft, IBM e Oracle, desenvolvem soluções com licenças comerciais. Também existem ferramentas *open source* com características similares, dentre as quais pode-se citar o PostgreSQL, como um SGBD completo que está entre os mais usados atualmente, e o SpagoBI que possui diversos componentes para realização do ETL e exploração dos dados.

Segundo *The PostgreSQL Global Development Group (2016)*, o PostgreSQL começou a ser desenvolvido em 1977. É, portanto, um banco maduro, que possui as principais características desejadas em um banco de dados, como, por exemplo, recuperação automática após falhas do sistema, controle de concorrência, suporte a transações, Backup *on-line*, tamanho ilimitado de registro (os únicos limites de armazenamento de tipos de dados são impostos pela configuração do hardware), etc. O PostgreSQL é multiplataforma, e oferece baixo custo total de instalação, reduzindo de forma significativa os custos de administração, suporte e licenciamento e, ao mesmo tempo, fornece alta performance, confiabilidade e escalabilidade (The PostgreSQL Global Development Group, 2016).

Segundo *Engineering Ingegneria Informatica S.p.A. (2016)*, o SpagoBI é uma plataforma integrada que abrange todos os componentes de uma solução de *Business Intelligence*, tanto em termos de análise e gerenciamento de dados, administração e segurança. Oferece soluções para relatórios, análise multidimensional (OLAP), *data mining*, etc. Também possui ferramentas para *Extract, Transform, and Load*, auxiliando o administrador na concepção e manutenção da solução. O SpagoBI permite a integração de soluções proprietárias para construir uma plataforma adequada para um problema particular. Ele tem uma estrutura modular em que todos os módulos estão relacionadas com o sistema de núcleo.

## **2.7 Desafios e Benefícios**

Para Turban et al. (2009), o principal desafio da implementação de uma solução de *Business Intelligence* está na qualidade dos dados. Segundo Orr (1998 apud BRITTO; JÚNIOR, 2006), qualidade de dados é o grau de aderência entre as visões apresentadas pelos dados armazenados e os mesmos dados no mundo real. Porém, em sistemas de grande porte a questão não é garantir que a aderência dos dados ao mundo real seja de 100%, ou seja, que a qualidade de dados seja perfeita, mas sim que seja suficientemente precisa, atualizada e consistente para que a organização possa sobreviver e tomar decisões razoáveis. Sendo assim, A qualidade de dados também pode ser definida como o grau de adequação ao uso, fazendo com que o termo ganhe caráter relativo, pois dados considerados de qualidade apropriada para um uso específico podem ser insuficientes para outro.

Muitas empresas não tomam o cuidado que deveriam na modelagem de seus sistemas de armazenamento de dados, o que acarreta em sistemas que geram dados de baixa qualidade. Para Turban et al. (2009), essa prática faz com que o processo de ETL seja mais complexo. Do contrário, quando os dados apresentam alta qualidade, se tornam um ativo da empresa, proporcionando um processo de ETL mais simples, resultando em economia com instalação e manutenção, além de maior eficiência no processo de extração de conhecimento.

Ao se implementar um sistema de BI, são trazidos benefícios para a instituição. Segundo Chaudhuri et al. (2011), atualmente é difícil encontrar uma empresa de sucesso que não aplique nenhuma técnica de BI nos seus negócios. Entretanto, mensurar o valor do *Business Intelligence* para o negócio é difícil, pois é preciso levar em conta alguns benefícios como o fato do *Data Warehouse* ser uma infraestrutura que serve também para outras aplicações da empresa (TURBAN et al., 2009 apud LIMA; LIMA, 2010).

Segundo Botelho e Filho (2014), as empresas que não utilizam BI podem estar deixando de obter informações sobre o seu negócio que as leve a tomar melhores decisões. Algumas, porque não conseguem visualizar os benefícios que tais ferramentas podem representar no seu processo decisório; outras, porque possuindo o potencial analítico, não utilizam as ferramentas adequadas para análises, e conseqüentemente produzem resultados insuficientes.

## 2.8 Trabalhos Relacionados

Esta seção apresenta alguns trabalhos relacionados estudados para o desenvolvimento deste trabalho. Foi encontrado na literatura apenas um relato tratando especificamente da aplicação de *Business Intelligence* em bibliotecas universitárias. Por isso, são mostradas também abordagens de aplicação destes recursos como apoio à qualidade de ensino e gestão universitária.

Especificamente sobre aplicação de técnicas de *Business Intelligence* em bibliotecas, Zucca (2013) descreve uma justificativa para o desenvolvimento de uma infraestrutura de BI, na Universidade da Pensilvânia, chamada MetriDoc, capaz de facilitar a coleta, integração e apresentação de dados das atividades de bibliotecas. A publicação tem por objetivo buscar parcerias para continuar o desenvolvimento do sistema. O autor informa que a infra-estrutura estava sendo implementada em fase de teste em mais duas universidades, a do estado da Carolina do Norte e de Chicago, e não

apresenta resultados.

Não foram encontrados na literatura registros de aplicação de técnicas de BI em bibliotecas de instituições de ensino brasileiras. Porém, Ciupak et al. (2013) realizaram um estudo onde fazem uma análise do uso de tecnologias de *Business Intelligence* como facilitadoras da gestão universitária no contexto da Universidade Estadual do Oeste do Paraná (UNIOESTE). A pesquisa visou verificar se a implementação de OLAP na UNIOESTE facilitaria a extração de informações das suas bases de dados de maneira mais amigável, conferindo mais autonomia a seus usuários. Posteriormente, foi feita uma avaliação da solução utilizando questionários, para mensurar a satisfação do usuário, onde a média atribuída pelos participantes foi 91,5, em uma escala de 0 a 100, perfazendo assim uma avaliação bastante positiva.

Reis et al. (2010) apresentam um estudo que investiga o uso de *Business Intelligence* como suporte para a extração do conhecimento de bancos de dados para a formulação de estratégias, com o intuito de buscar a melhoria da qualidade de ensino nas escolas ligadas à Secretaria de Estado da Educação e Inovação do Estado de Santa Catarina. Este estudo ainda analisa a importância da gestão de conhecimento na geração de informações de suporte à tomada de decisões e como forma de auxílio na gestão estratégica. O processo mostrou ser possível extrair conhecimento relevante mediante mineração de dados, gerando conhecimentos que podem ser interpretados e avaliados.

Outro trabalho destacado é o de Almeida e Camargo (2015), realizado na própria Universidade Federal do Pampa. Este trabalho apresenta uma abordagem de implementação de *Business Intelligence* sobre dados de desempenho acadêmico com o objetivo de otimizar a aplicação de seus recursos. Ao final deste trabalho, foi possível a criação de *dashboards* com informações de matrículas, ingressos formados e evasão.

Com base principalmente nestes quatro trabalhos e na revisão de literatura apresentada, foi realizado o projeto da solução de BI que está descrito no capítulo 3.

### 3 SOLUÇÃO PROPOSTA

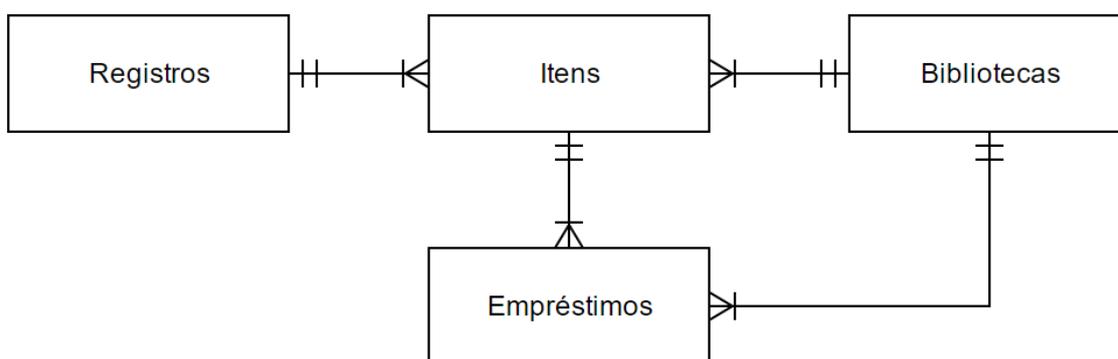
Este capítulo apresenta a solução de *Business Intelligence* proposta para auxiliar a gestão das bibliotecas da UNIPAMPA. Nele são apresentados os aspectos de projeto, tais como: definição dos indicadores de desempenho, criação do *Data Warehouse*, processo de ETL, exploração e visualização dos dados e também aspectos da implantação da solução.

#### 3.1 Obtenção dos dados

A primeira etapa deste trabalho consistiu na solicitação de acesso para uma consulta aos dados das bibliotecas UNIPAMPA. A autorização para a coleta de dados foi concedida por meio da "carta de ciência sobre trabalho de conclusão de curso e aceite quanto a coleta de dados na universidade" apresentada no apêndice B.

Com isso, foram gerados arquivos no formato CSV, cujos dados puderam ser carregados em sua totalidade para um banco de dados fonte, criado no SGBD PostgreSQL e modelado com base no banco de dados da UNIPAMPA, apresentado na figura 6. Esse banco foi criado com o intuito de simular banco de dados da universidade, de forma que o processo de *Extract, Transform, and Load* já fosse preparado pensando na implantação da solução de BI na UNIPAMPA.

Figura 6 – Modelo conceitual do banco de dados das bibliotecas da UNIPAMPA



Fonte: Próprio Autor (2016)

O modelo é composto por quatro tabelas, sendo elas:

- Registro: Nela estão presentes dados referentes ao registro bibliográfico do livro, tais como: título, subtítulo, autor, ano de publicação, editora, número da edição,

etc;

- Item: Contém dados relativos a exemplares. Dentre eles, estão uma referência para o seu respectivo registro bibliográfico e para a biblioteca à qual pertence, além de dados como: número de patrimônio; data, hora e valor de aquisição; registro da estante na qual se encontra, se está disponível apenas para consulta local, etc;
- Biblioteca: Guarda dados referentes às bibliotecas, tais como: nome, sigla, localização, telefone, email, etc;
- Empréstimo: Inclui referência para o item que está sendo emprestado, a biblioteca onde está acontecendo o evento, pessoa que está retirando o item, funcionário que está realizando a operação, além de dados como data e hora da ocorrência.

### 3.2 Identificação dos Indicadores de Desempenho

Para que apenas os dados relevantes para exploração fossem processados no ETL e carregados no DW, foi realizada a definição dos indicadores de desempenho. A principal atividade em uma biblioteca está no apoio ao ensino, pesquisa e extensão por meio do empréstimo do seu acervo de livros. Sendo assim, é possível dizer que o principal indicador de desempenho de uma biblioteca é a quantidade de empréstimos que ela realiza.

Porém, valores totais de empréstimos são insuficientes para medir o desempenho pois, o que para uma biblioteca com poucos exemplares pode ser considerado um valor alto, para outra maior, pode ser baixo. Por isso, é importante relacionar os empréstimos com outros dados capazes de trazer significado à eles.

Então, com base nos dados disponíveis para análise e no levantamento de requisitos foram definidos os seguintes indicadores de desempenho, baseados em análises do acervo e sua movimentação:

- Quantidade de empréstimos por *campus*, filtrados por mês e ano;
- Quantidade de empréstimos por ano, filtrados por *campus* e mês;
- Quantidade de empréstimos por mês, filtrados por *campus* e ano;
- Quantidade de empréstimos por horário, filtrados por *campus*, mês e ano;
- Quantidade de exemplares por *campus*;
- Custo do acervo por *campus*;

- Relatórios de exemplares mais emprestados;
- Relatórios de exemplares nunca emprestados;
- Quantidade de exemplares nunca emprestados por *campus*;
- Custo do acervo nunca emprestado;
- Índice de empréstimos por exemplar por *campus*;
- Índice de empréstimos por exemplar por título.

### 3.3 Projeto do *Data Warehouse*

O projeto do *Data Warehouse* foi realizado de forma que nele fossem colocados todos os dados necessários para o monitoramento dos indicadores. Sua implementação foi baseada na metodologia *bottom-up* proposta por Kimball e Ross (2013), ou seja, primeiramente foram definidos os *Data Marts*, para que depois eles fossem reunidos formando um DW completo.

Para compor o *Data Warehouse*, foram criados dois *Data Marts* multidimensionais. O primeiro, mostrado na figura 7, foi desenvolvido para a exploração de dados relacionados aos empréstimos e para o cruzamento com informações de data, hora e campus. O segundo, mostrado na figura 8, foi projetado para dar suporte às análises relacionadas ao acervo de livros, tais como: quantidade, custos, exemplares mais emprestados, etc.

Figura 7 – *Data Mart* de empréstimos

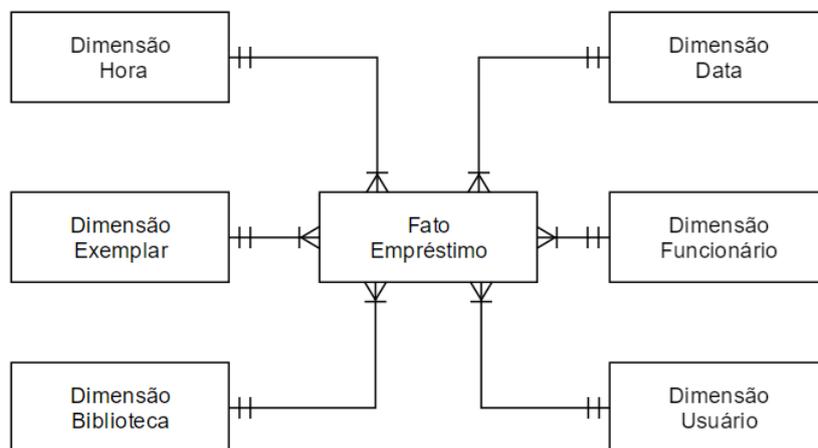
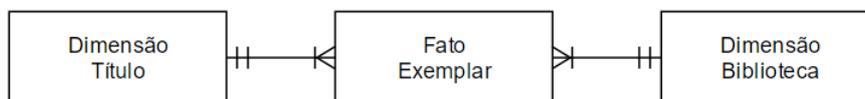
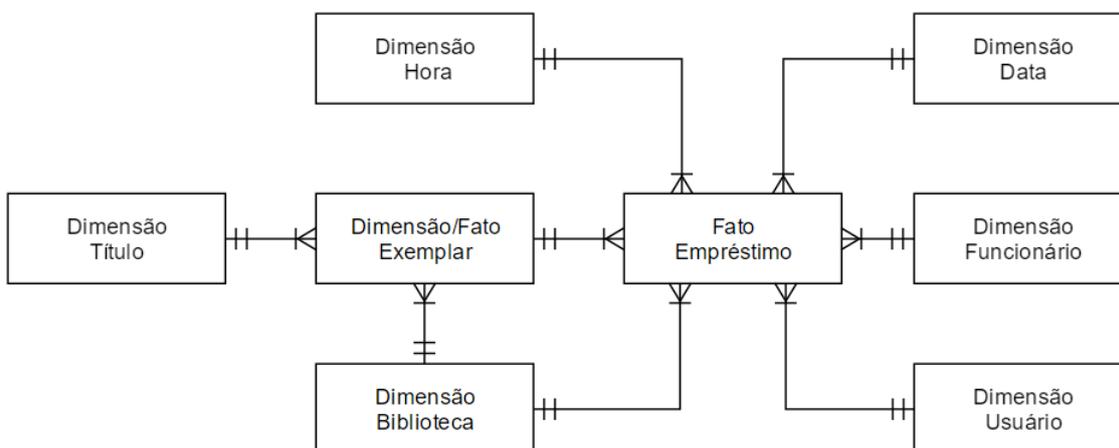


Figura 8 – *Data Mart* de exemplares

Fonte: Próprio Autor (2016)

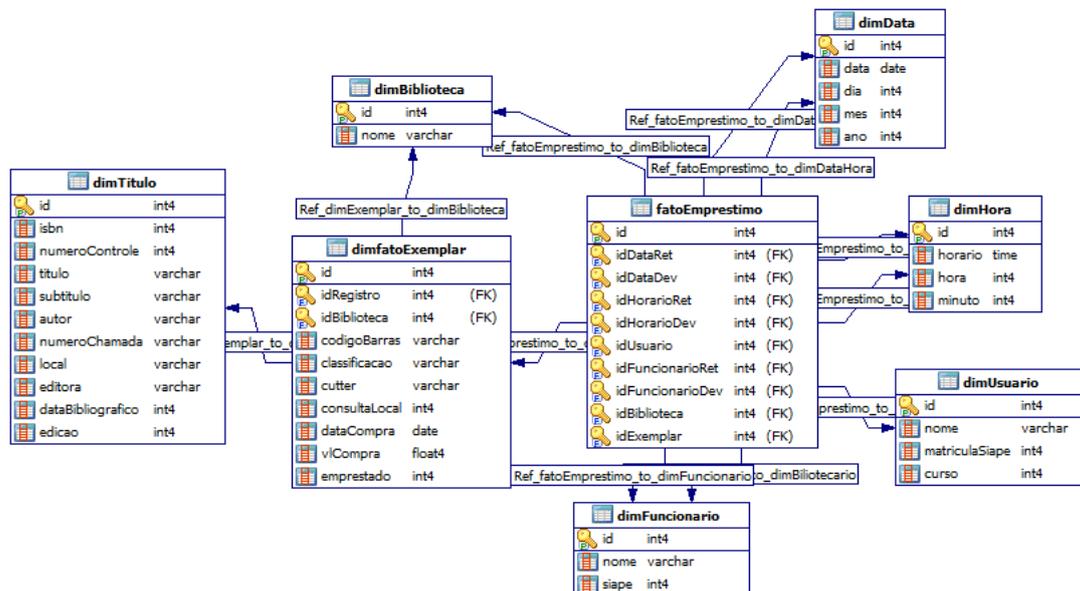
A união de ambos deu origem ao esquema principal do *Data Warehouse*, cujo modelo conceitual é apresentado na figura 9. A tabela "Dimensão/Fato Exemplar" recebeu esse nome pois em um DM se comporta com dimensão do fato empréstimo e no outro é o fato de interesse da análise.

Figura 9 – Modelo conceitual do *Data Warehouse*

Fonte: Próprio Autor (2016)

O modelo lógico desenvolvido é apresentado na figura 10. Nele foram criados campos para guardar o ISBN, e também dados de usuários e funcionários das bibliotecas da UNIPAMPA. Pois, apesar de estes dados não estarem disponíveis durante o projeto da solução, o DW já fica preparado para o caso de eles serem incluídos futuramente.

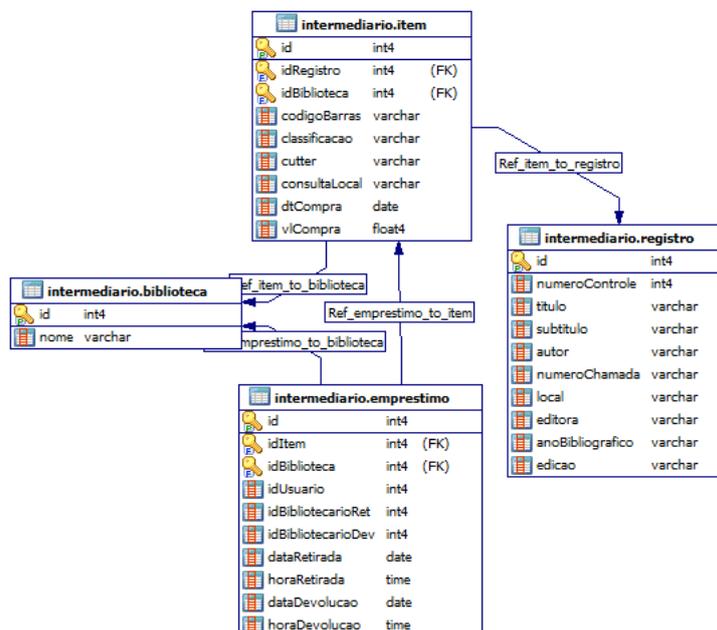
Figura 10 – Modelo lógico do esquema principal do DW



Fonte: Próprio Autor (2016)

No *Data Warehouse* foi desenvolvido também, um esquema intermediário para servir como área de preparação durante o processo de ETL. Esse esquema possui modelo correspondente ao banco de dados de origem, porém guarda só aqueles campos relevantes para análise e necessários para povoar o DW. O modelo lógico do esquema intermediário é apresentado na figura 11.

Figura 11 – Modelo lógico do esquema intermediário do DW



Fonte: Próprio Autor (2016)

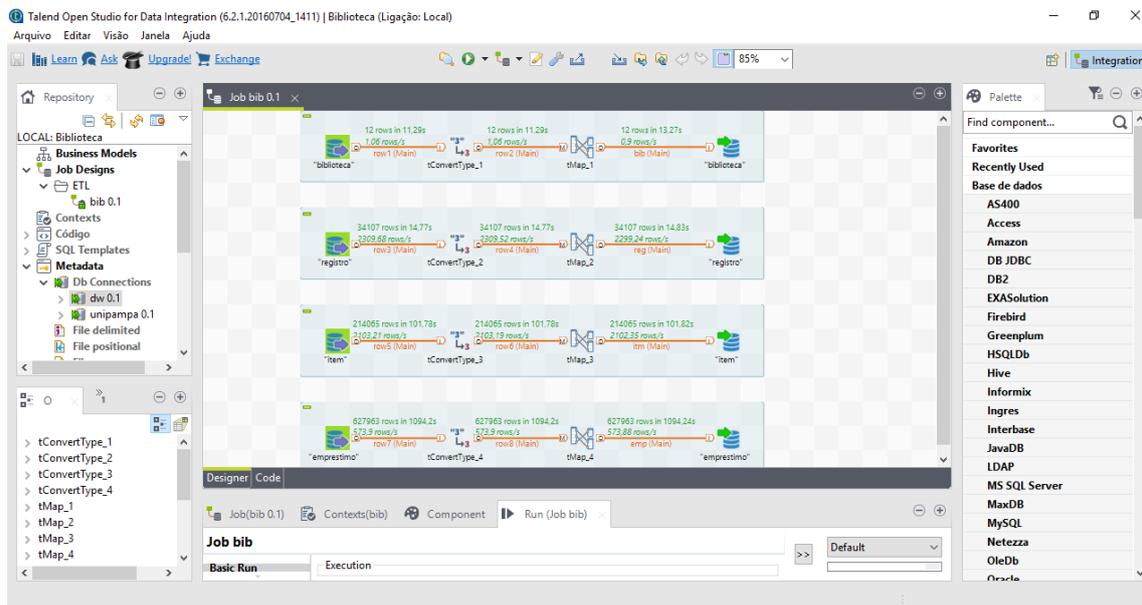
### 3.4 Processo de *Extract, Transform, and Load*

O processo de ETL é composto por três etapas, sendo que a primeira delas, a de extração, foi realizada por meio da ferramenta Talend Data Integration<sup>1</sup>. Essa ferramenta possui *drivers* para conexão com bancos de dados diversos, inclusive o DB2, que é o banco de dados proprietário da IBM, usado pela UNIPAMPA e, o PostgreSQL<sup>2</sup>, banco de dados *open source* usado no projeto do *Data Warehouse*. Além disso, o Talend Data Integration possui componentes de mapeamento e conversão de dados possibilitando assim, a extração de dados de uma base de dados, para carga em outra. A figura 12 mostra a interface dessa ferramenta após a execução da extração.

<sup>1</sup><<https://www.talend.com/products/talend-open-studio>>

<sup>2</sup><<https://www.postgresql.org>>

Figura 12 – Interface da Ferramenta Talend Data Integration



Fonte: Próprio Autor (2016)

O Talend Data Integration gera uma aplicação em Java, que é útil para a automatização do processo de ETL, pois possibilita que sua execução seja agendada no próprio sistema operacional onde a ferramenta é executada, sem a necessidade de usar a interface. Essa aplicação fica então responsável por extrair os dados da fonte original e mapeá-los para o esquema intermediário do *Data Warehouse*.

Para as etapas de transformação e carga dos dados para o esquema principal do DW, foram criados *triggers* (gatilhos) em cada uma das tabelas do esquema intermediário. Esses gatilhos são responsáveis por executar de forma automática uma função para cada registro inserido nas tabelas do esquema intermediário. As funções foram desenvolvidas com PL/pgSQL, e são responsáveis por tratar inconsistências dos dados, colocá-los em formato compatível e carregá-los no esquema principal do DW.

A PL/pgSQL é uma linguagem estendida da SQL que tem por objetivo auxiliar as tarefas de programação no PostgreSQL por meio de recursos que permitem o controle de fluxo de programas, tais como *loops* estruturados e controle de decisão.

O código desenvolvido é apresentado no apêndice C. Nele existem quatro funções, sendo elas:

- *sincronizaBib*: Carrega os dados da tabela "biblioteca" do esquema intermediário, para a tabela "dimBiblioteca" do esquema principal, não sendo necessárias transformações.

- **sincronizaReg:** Carrega os dados da tabela "registro" do esquema intermediário, para a tabela "dimTitulo" do esquema principal. Os campos "anoBibliografico" e "edição", que são *strings* no banco de dados da UNIPAMPA, são convertidos para inteiros. Essa transformação é necessária pois na coluna "anoBibliografico" um mesmo ano aparece de diversas formas. Por exemplo, o ano de 2016 pode aparecer como "2016", "2016.", "c2016.", etc. sendo que os caracteres não inteiros não agregam qualquer valor aos dados, pelo contrário, se for realizado um agrupamento com base nesse dado, a ferramenta não irá os reconhecer como iguais, causando inconsistência na informação apresentada. O mesmo acontece no campo "edição", onde esta também pode aparecer de diferentes formas, por exemplo, "3", "3ed", "3.ed", "3 ed.", etc.
- **sincronizaItm:** Carrega os dados da tabela "item", do esquema intermediário, para a tabela "dimfatoExemplar" do esquema principal. O campo "consultaLocal", que é uma *string* no banco de dados original, é convertido para inteiro, sendo atribuído o valor "1" quando o exemplar é disponível apenas para consulta local, e o valor "0" quando não é. Essa transformação também é necessária devido ao fato de o dado aparecer de diferentes formas, como por exemplo, "consulta local", "CONSULTA LOCAL", "Consulta Local", etc. e, em alguns casos, também com erros de digitação.
- **sincronizaEmp:** Carrega os dados da tabela "emprestimo", do esquema intermediário, para as tabelas "dimfatoExemplar", "dimFuncionario", "dimUsuario", "dimData", "dimHora" e "fatoEmprestimo" do esquema principal. Na tabela "dimfatoExemplar", a cada inserção, é feito um incremento no campo "emprestado" do item correspondente. Nas tabelas "dimFuncionario" e "dimUsuario" é feita uma consulta a cada inserção, para verificar se o registro pode ser inserido ou não, caso já esteja cadastrado. Nessas tabelas são mantidas as chaves primárias originais. Processo semelhante ocorre nas tabelas "dimData" e "dimHora", com a diferença de que as consultas ou inserções devem retornar o valor de chave primária registrado, uma vez que este é controlado pelo banco. Nos dados de horário, é feito um arredondamento para que os dados possuam precisão de 15 minutos, e não mais de segundos. Isso é feito para facilitar a visualização e interpretação de informações. Por fim, a chave primária original e todas as estrangeiras registradas são gravadas na tabela "fatoEmprestimo".

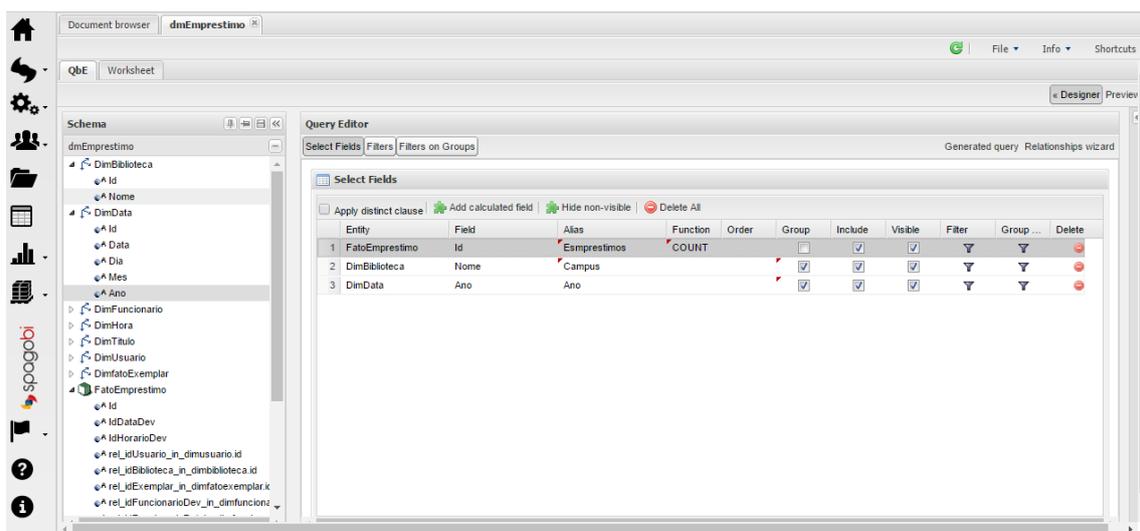
Ao final, têm-se um processo automatizado para realização do ETL, e assim, pode

dar-se início à exploração.

### 3.5 Exploração dos Dados

A exploração dos dados foi feita com o SpagoBI<sup>3</sup>. Para isso, inicialmente foi necessária a realização de algumas configurações na plataforma. A primeira, consiste na conexão da ferramenta com o DW, e do SpagoBI Studio com o SpagoBI Server. Depois, no SpagoBI Studio são definidos os DMs a serem explorados no Server. Após a criação dos modelos no Studio, eles são carregados no SpagoBI Server, para que sejam definidos os *datasets*. Os *datasets* são definidos por meio de consultas SQL nos DMs. O SpagoBI Server possui um módulo chamado *Query by Example* (QbE) que facilita a criação das consultas. A figura 13 mostra a interface desse módulo.

Figura 13 – Interface do módulo QbE



Fonte: Próprio Autor (2016)

Com o QbE é possível montar consultas SQL arrastando os campos desejados para o *Query Editor*, definindo filtros, agrupamentos e funções de agregação. Essa interface proporciona que mesmo usuários com pouco conhecimento da linguagem SQL possam definir consultas personalizadas e também é útil para usuários avançados pois torna o processo mais ágil, sem a necessidade de escrita de códigos.

Com este recurso, foram definidos os *datasets* para a criação de gráficos e relatórios para o monitoramento de cada um dos KPIs. Os resultados são apresentados na

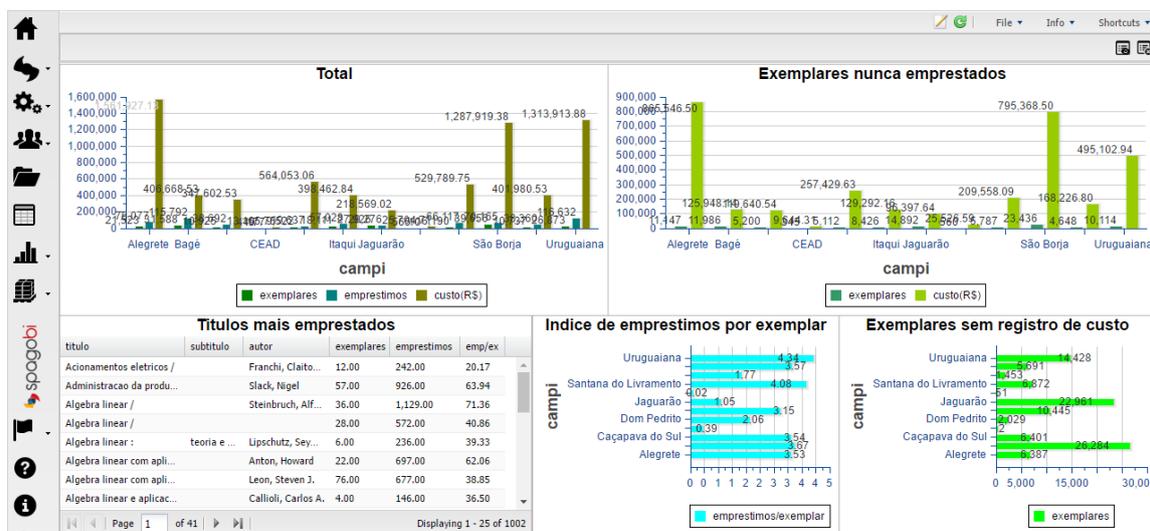
<sup>3</sup><<https://www.spagobi.org>>

seção 3.6.

### 3.6 Apresentação dos Conhecimentos Extraídos

Para a apresentação dos conhecimentos extraídos dos dados, o SpagoBI Server oferece um módulo para criação de documentos no formato *cockpit*. Os *cockpits* são *dashboards* interativos, ou seja, um painel de indicadores para o monitoramento dos KPIs. Com esse recurso, foi desenvolvido um *dashboard* para o monitoramento do total de exemplares, empréstimos e custo do acervo por biblioteca; quantidade e custo dos exemplares que não são "consulta local" e nunca foram emprestados; títulos mais emprestados e índice de "empréstimos por exemplar" e nunca foram emprestados; títulos mais emprestados e índice de "empréstimos por exemplar" por título e *campus*. Esse painel de indicadores é mostrado na figura 15.

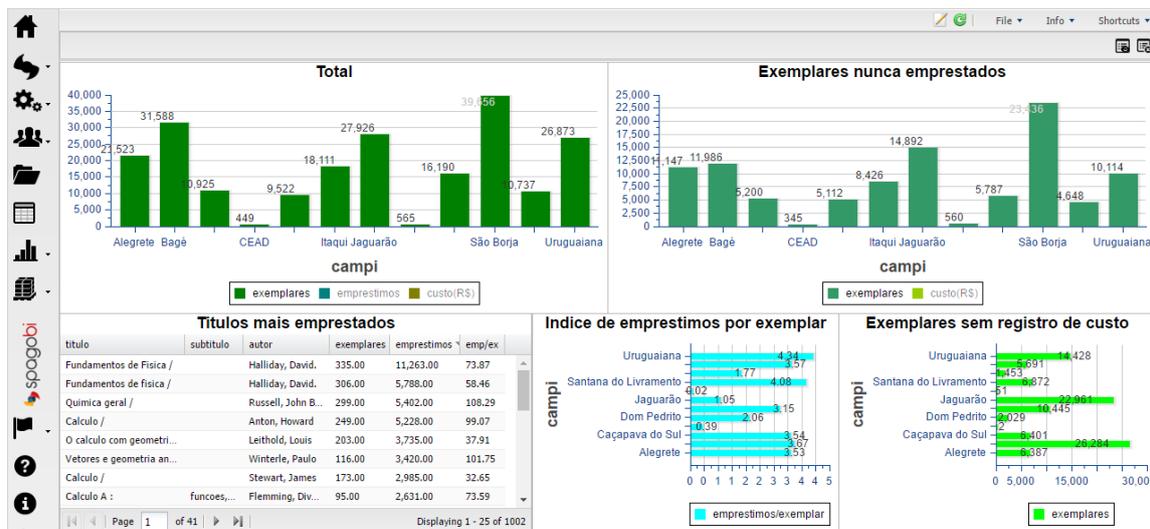
Figura 14 – Dashboard de acervo completo



Fonte: Próprio Autor (2016)

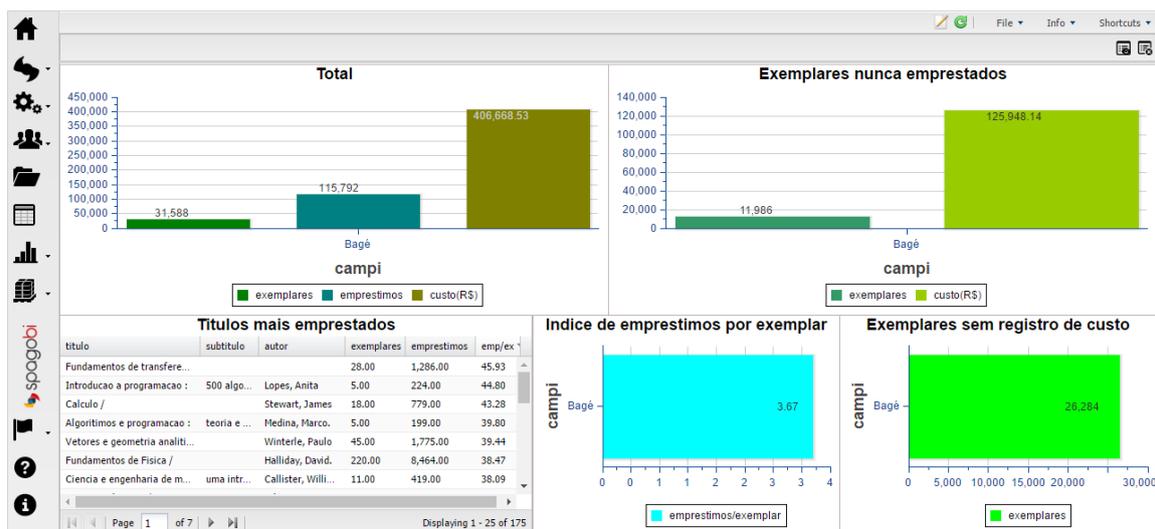
Esse *dashboard* é interativo pois possibilita que certas medidas sejam ocultadas ou mostradas com um clique na legenda e as linhas contidas na tabela podem ser ordenadas com base no número de exemplares, empréstimos ou do índice empréstimos por exemplar. A figura 15 mostra o painel definido para apresentar apenas as informações referentes à quantidade de exemplares no gráfico "Total" e "Exemplares nunca emprestados", com os títulos ordenados pelo maior número de empréstimos.

Figura 15 – Dashboard de acervo mostrando a quantidade de exemplares



Fonte: Próprio Autor (2016)

Além disso, clicando-se sobre uma medida de um determinado *campus*, todos os componentes do painel são filtrados mostrando apenas as informações referentes àquele *campus*. A figura 16 mostra o painel definido para apresentar informações referentes ao *campus* Bagé, com os títulos ordenados pelo índice empréstimos por exemplar.

Figura 16 – Dashboard de acervo filtrado para o *campus* Bagé

Fonte: Próprio Autor (2016)

Outra possibilidade de exibição do painel é mostrada na figura 17, e se refere ao custo total do acervo e dos exemplares que não são "consulta local" e nunca foram

emprestados. Porém, esses números são inconsistentes com o número de exemplares, e para perceber isso pode-se analisar por exemplo, que apesar de o tamanho do acervo de Bagé ser aproximadamente 50% maior que o de Alegrete, o seu custo é quase 300% menor. Isso é explicado no gráfico que relaciona os exemplares sem custo registrado, onde é possível verificar que em Bagé mais de 26.000 exemplares estão sem o valor de compra registrado, o que representa 83% do total. Já em Alegrete, são 6.387 exemplares sem o registro, representando cerca de 30%.

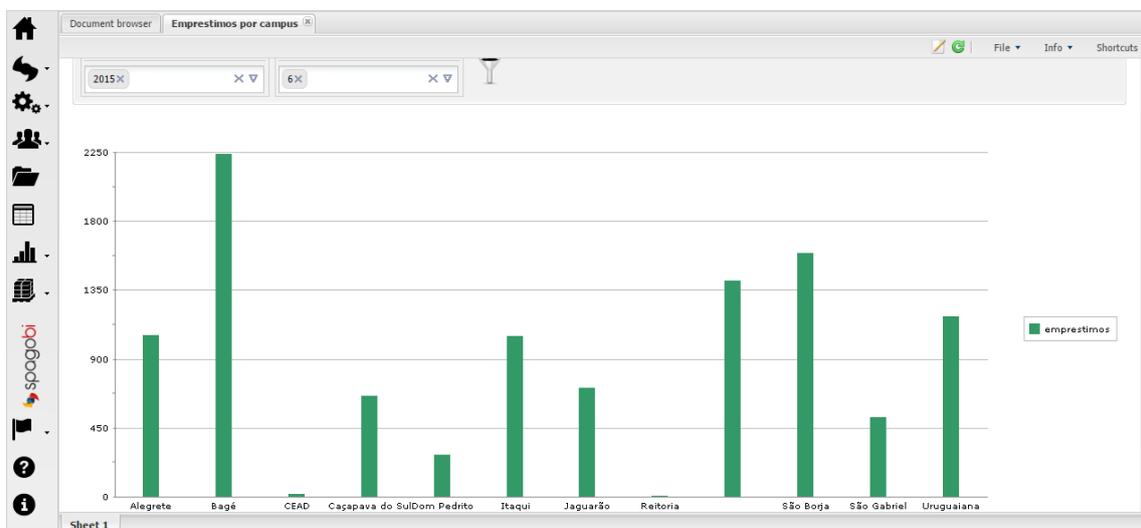
Figura 17 – Dashboard de acervo mostrando o custo dos exemplares



Fonte: Próprio Autor (2016)

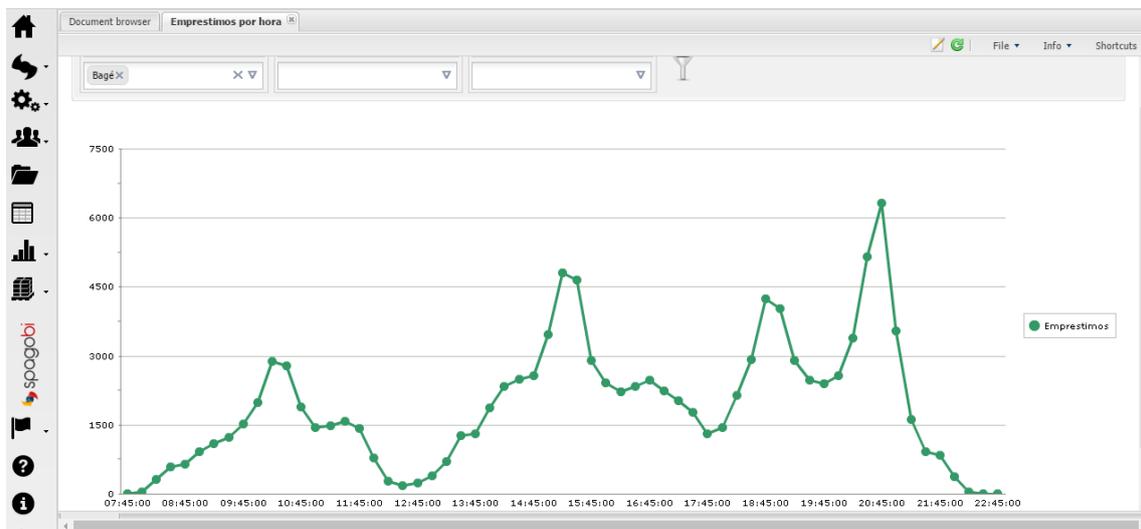
O módulo que cria os *cockpits* interativos traz diversos recursos que permitem a análise, cruzamento e comparação de diversas medidas de desempenho simultaneamente, mas não se mostra tão eficiente à medida que cresce o número de dimensões relacionadas entre os componentes. Isso ocorre devido ao aumento da complexidade das consultas para atualização dos *datasets*, que combinado com a necessidade de atualização de vários *datasets* simultaneamente, causa lentidão no processo, o que vai contra o requisito NF003 do projeto apresentado no apêndice A.

Por isso, para monitorar os indicadores que envolvem vários filtros, foram criados gráficos e relatórios individuais em forma de documentos chamados de *ad hoc* (específicos) no SpagoBI. Um exemplo desse tipo de documento é apresentado na figura 18, que mostra a quantidade de empréstimos por *campus* filtrados por mês e ano.

Figura 18 – Empréstimos por *campus* no ano de 2015

Fonte: Próprio Autor (2016)

Outro conhecimento extraído durante a exploração são os horários com maior número de empréstimos nas bibliotecas da UNIPAMPA. Essa análise é relevante, principalmente quando individualiza por *campus*, uma vez que permite compreender o perfil dos usuários. Pode-se definir também filtros referentes ao mês e ano, para avaliar esse comportamento em períodos específicos. A figura 19 apresenta esse conhecimento referente à biblioteca do *campus* Bagé. Nela é possível observar picos de empréstimos nos horários de intervalo dos turnos de aulas, e também em meio aos turnos da tarde e noite.

Figura 19 – Empréstimos por horário no *campus* Bagé

Fonte: Próprio Autor (2016)

Também foram criados relatórios no formato de arquivos *ad hoc*. Assim como nos gráficos, é possível definir filtros para os dados de acordo com as necessidades. A figura 20 apresenta um relatório de exemplares nunca emprestados filtrados para o *campus* Bagé.

Figura 20 – Relatório de exemplares do *campus* Bagé nunca emprestados

captura de tela de um relatório de exemplares nunca emprestados no campus Bagé. O relatório é exibido em uma interface de usuário com uma barra de filtros e uma tabela de dados.

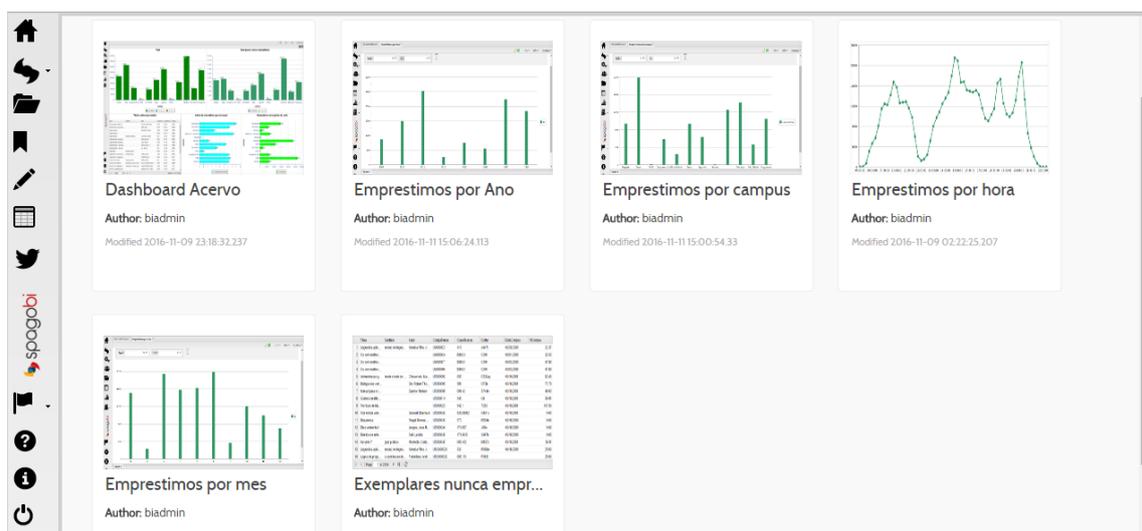
Título	Subtítulo	Autor	CodigoBarras	Classificacao	Cutter	DataCompra	VCCompra	
1	Linguística aplic...	ensino de lingua...	Almeida Filho, J...	UBG000026	530	R585m	04/18/2008	25.00
2	Logica de progr...	a construcao de...	Forbellone, Andr...	UBG000032	005.115	F692l		25.00
3	Logica de progr...	a construcao de...	Forbellone, Andr...	UBG000033	005.115	F692l		25.00
4	A aventura do li...	do leitor ao nave...	Chartier, Roger.	UBG000045	002	C486a	03/10/2008	53.00
5	A aventura do li...	do leitor ao nave...	Chartier, Roger.	UBG000046	002	C486a	03/10/2008	53.00
6	A aventura do li...	do leitor ao nave...	Chartier, Roger.	UBG000047	002	C486a	03/10/2008	53.00
7	A aventura do li...	do leitor ao nave...	Chartier, Roger.	UBG000048	002	C486a	03/10/2008	53.00
8	A aventura do li...	do leitor ao nave...	Chartier, Roger.	UBG000049	002	C486a	03/10/2008	53.00
9	A aventura do li...	do leitor ao nave...	Chartier, Roger.	UBG000050	002	C486a	03/10/2008	53.00
10	A aventura do li...	do leitor ao nave...	Chartier, Roger.	UBG000051	002	C486a	03/10/2008	53.00
11	A aventura do li...	do leitor ao nave...	Chartier, Roger.	UBG000052	002	C486a	03/10/2008	53.00
12	Quimica Organi...		Bruice, Paula Yu...	UBG000068	547	B886q	03/10/2008	100.00
13	Quimica Organi...		Bruice, Paula Yu...	UBG000069	547	B886q	03/10/2008	100.00
14	Boas Maneiras ...	um guia pratico ...	Ribeiro, Celia.	UBG000079	395.52	R484b	03/10/2008	30.00
15	Os persas /		Esquilo	UBG000206	880	E77p		37.00
16	Os persas /		Esquilo	UBG000207	880	E77p		37.00

Fonte: Próprio Autor (2016)

Os documentos, nos formatos *cockpit* e *ad hoc* gerados, podem ser colocados na área de análise do usuário, que também funciona como um *dashboard*, a medida que

reúne outros documentos *ad hoc* e até *dashboards*. Essa área é mostrada na figura 21.

Figura 21 – Área de análise

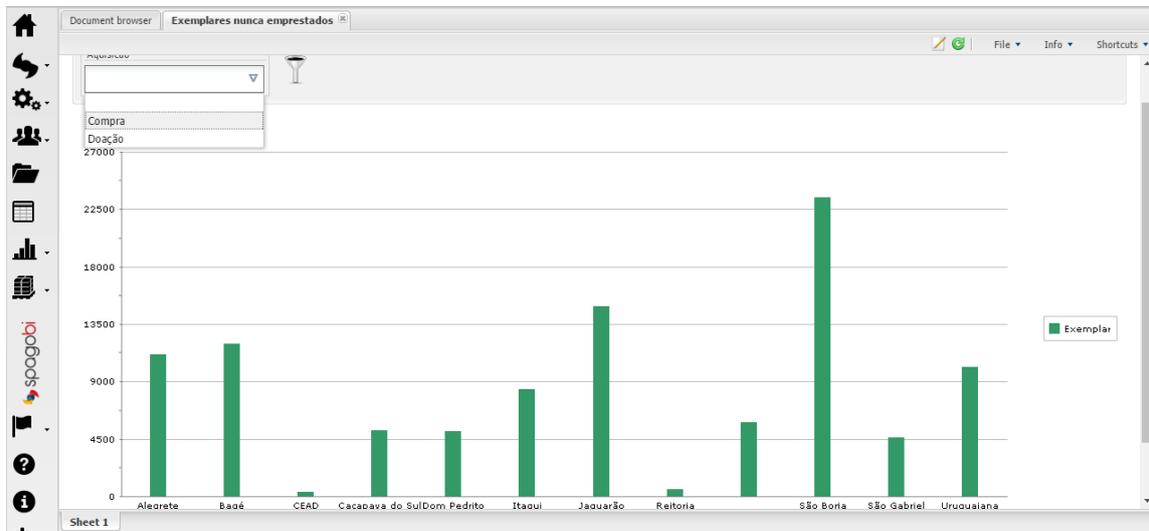


Fonte: Próprio Autor (2016)

Para demonstrar o funcionamento e as funcionalidades da solução de Business Intelligence desenvolvida foi desenvolvido um vídeo que pode ser acessado em <<https://youtu.be/QwGXJ25-AuY>><sup>4</sup>. Todas as interfaces da seção 3.6 são apresentadas em tamanho maior no apêndice D para uma melhor visualização dos detalhes.

Concluído o projeto da solução de BI, foi realizada uma reunião com os gestores das bibliotecas da UNIPAMPA com o objetivo de apresentar os resultados obtidos, para que os usuários pudessem opinar e dar sugestões de melhorias para a solução. Com base nisso, foi criado mais um gráfico. Esse gráfico é apresentado na figura 22 e mostra a quantidade de exemplares por *campus*, com a opção de filtragem por forma de aquisição.

<sup>4</sup>Para obter acesso ao vídeo, solicite permissão ao autor através do email: leogschulte@gmail.com

Figura 22 – Exemplos nunca emprestados por *campus*

Fonte: Próprio Autor (2016)

## 4 CONSIDERAÇÕES FINAIS

Após a conclusão do projeto da solução é possível considerar que a metodologia adotada obteve sucesso, já que puderam ser criados gráficos, relatórios e *dashboards* que mostram várias informações relevantes. A reunião com os gestores, para apresentação dos resultados, indicou que a solução pode auxiliar no processo decisório nas bibliotecas da UNIPAMPA, já que os usuários demonstraram satisfação e sugeriram a criação de gráficos que não haviam sido citados no levantamento de requisitos, mostrando que passaram a ver ainda mais potencial de descoberta de conhecimento na solução de BI.

Porém, é importante salientar que os resultados poderiam ser melhores se não existissem problemas de qualidade nos dados explorados. Esses problemas são causados basicamente pela ausência de dados relevantes para análise, e falta de validação dos registros na entrada. Por isso, dos requisitos funcionais, apenas o RF005 *campi* com menor número de empréstimos, RF006 *campi* com maior número de empréstimos, RF007 horários com maior número de empréstimos e RF008 exemplares sem registro do valor de compra, foram cumpridos totalmente, pois para todos os outros houve ausência de algum dado relevante. Pode-se dizer que os requisitos funcionais RF003 exemplares nacionais nunca emprestados e RF004 exemplares importados nunca emprestados, foram contemplados parcialmente, visto que só não foi possível fazer a diferenciação entre títulos nacionais e importados. Esses problemas poderiam ser corrigidos caso o ISBN fosse adicionado aos dados, pois com base nele pode-se recuperar informações de título, autor, idioma, país, edição, editora, etc. Em contrapartida, a definição dos KPIs indicou dados com potencial de exploração que durante o levantamento de requisitos não foram citados. Os requisitos não funcionais foram totalmente contemplados pela solução de BI desenvolvida.

Como trabalhos futuros, serão realizadas a implantação e validação da solução. Para a implantação, será necessária a autorização do Núcleo de Tecnologia da Informação e Comunicação (NTIC) da UNIPAMPA para realizar a instalação das ferramentas necessárias. Depois, será necessário instruir os usuários, para que possam utilizar de forma adequada a solução de BI desenvolvida. Para avaliar a solução, foi desenvolvido um questionário com base nos trabalhos realizados por Silva (2008) e Pimentel et al. (2012), que sugerem uma abordagem do *Technology Acceptance Model*, onde através da verificação da percepção do usuário a cerca da utilidade e facilidade de uso da tecnologia implantada, pode-se inferir a intenção de utilização. O questionário é apresentado no

apêndice E, e será submetido aos usuários do sistema após a implantação.

Além disso, seria interessante a ampliação dos dados explorados, com a inclusão de dados de movimentação de livros por curso, período do curso, categoria, área do conhecimento, etc. Também poderiam ser desenvolvidas soluções para eliminar, ou ao menos reduzir, os problemas de qualidade de dados existentes na UNIPAMPA. Essas soluções certamente incluem a conscientização dos usuários dos sistemas da UNIPAMPA para a importância da inclusão dos dados, e orientação para que essa inclusão seja feita de forma correta.

## REFERÊNCIAS

- ACKOFF, R. From data to wisdom. **Journal of Applied Systems Analysis**, v. 16, p. 3–9, 1989.
- ALMEIDA, A. M.; SCHULTE, L. G.; CAMARGO. **Aplicando técnicas de Business Intelligence sobre dados de desempenho acadêmico**. UNIPAMPA: Mostra de Iniciação Científica de Engenharias, 2015.
- ALMEIDA, A. M. R.; CAMARGO, S. Aplicando técnicas de business intelligence sobre dados de desempenho acadêmico: Um estudo de caso. **XI Escola Regional de Banco de Dados. Caxias do Sul-RS**, 2015.
- ANTONELLI, R. A. Conhecendo o business intelligence (bi). **Revista TECAP**, v. 3, n. 3, 2009.
- BARBIERI, C. **BI2: Business Intelligence: Modelagem e Qualidade**. Rio de Janeiro: Elsevier, 2011.
- BISPO, C. A. F. **Uma análise da nova geração de sistemas de apoio à decisão**. Thesis (PhD) — Universidade de São Paulo, 1998.
- BOTELHO, F. R.; FILHO, E. R. Conceituando o termo business intelligence: Origem e principais objetivos. **SISTEMAS, CIBERNÉTICA E INFORMÁTICA**, v. 11, n. 1, 2014.
- BRITTO, M. d. S.; JÚNIOR, J. R. d. A. Qualidade de dados para data warehouse roteiro de implementação. In: **International Conference on Information Systems and Technology Management**. [S.l.: s.n.], 2006. v. 3.
- CANALTECH. **Repositório digital do canaltech**. 2016. Available from Internet: <<http://imagens.canaltech.com.br/50682.69608-Data-Warehouse.png>>.
- CHAUDHURI, S.; DAYAL, U. An overview of data warehousing and olap technology. **ACM Sigmod record**, ACM, v. 26, n. 1, 1997.
- CHAUDHURI, S.; DAYAL, U.; NARASAYYA, V. An overview of business intelligence technology. **Communications of the ACM**, ACM, v. 54, n. 8, p. 88–98, 2011.
- CIUPAK, L. F.; BOSCARIOLI, C.; CATARINO, M. E. Análise do uso de tecnologias de business intelligence como facilitadoras à gestão universitária. **Brazilian Journal of Information Science**, Universidade Estadual Paulista, v. 7, n. 1, p. 47–69, 2013.
- COKINS, G. Why do large, once-successful companies fail? **Analytics Magazine**, May/June, 2012.
- COSTA, S. A. R. da. **Sistema de business intelligence como suporte à gestão estratégica**. Universidade do Minho - Escola de engenharia: Tese de Mestrado em Engenharia e Gestão de Sistemas de Informação, 2012.
- DATE, C. **Introduction to database systems**. 8ed. ed. [S.l.]: Pearson, 2004.

ELMASRI, S. B. N. R. **Fundamentals of Database Systems**. 4. ed. [S.l.]: Pearson Education, 2003.

Engineering Ingegneria Informatica S.p.A. . **spagobi | 100source Business Intelligence**. 2016. Available from Internet: <<http://www.spagobi.org>>.

GALASSI, C. R.; FÓRMICE, C. R.; DIONYSIO, R. C. C. **Modelagem De Banco De Dados**. Taquaritinga: AgBook, 2013.

HEUSER, C. **Projeto de banco de dados : Volume 4 da Série Livros didáticos informática UFRGS**. [S.l.]: Bookman, 2009. (Livros didáticos informática UFRGS).

INMON, W. H. **Building the Data Warehouse**. 4th ed. ed. [S.l.]: Wiley, 2005.

JARDIM, E. S.; OLIVEIRA, M. V. A. de; MORAVIA, R. V. Diferença entre banco de dados relacional e banco de dados dimensional. **Revista Pensar Tecnologia**, v. 4, n. 2, 2015.

KIMBALL, R.; ROSS, M. **The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling**. 3. ed. [S.l.]: Wiley, 2013.

KIMBALL, R. et al. **The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence**. [S.l.]: Wiley, 2010.

LIMA, A. V.; LIMA, D. M. Business intelligence como ferramenta gerencial no suporte ao processo de business performance management-doi: 10.5102/un. gti. v1i1. 1201. **Universitas: Gestão e TI**, v. 1, n. 1, 2010.

ORR, K. Data quality and systems theory. **Communications of the ACM**, ACM, v. 41, n. 2, p. 66–71, 1998.

PARMENTER, D. **Key performance indicators: developing, implementing, and using winning KPIs**. Third edition. [S.l.]: John Wiley & Sons, 2015.

PEREIRA, M. J. L. d. B.; FONSECA, J. G. M. Faces da decisão: as mudanças de paradigmas e o poder da decisão. **Makron Books**, São Paulo, 1997.

PIMENTEL, V.; SOARES, J. et al. A utilização do computador na educação: aplicando o technology acceptance model (tam). Universidade Federal da Paraíba, 2012.

PRIMAK, F. V. **Decisões com B.I. (Business Intelligence)**. Rio de Janeiro: Editora Ciência Moderna, 2008.

REIS, E. S.; ANGELONI, M. T.; SERRA, F. R. Business intelligence como tecnologia de suporte a definição de estratégias para a melhoria da qualidade de ensino. **Informação & Sociedade**, Universidade Federal da Paraíba-Programa de Pós-Graduação em Ciência da Informação, v. 20, n. 3, 2010.

ROWLEY, J. E. The wisdom hierarchy: representations of the dikw hierarchy. **Journal of information science**, Sage Publications, 2007.

SEZÕES, C.; OLIVEIRA, J.; BAPTISTA, M. Business intelligence. **São João do Estoril, Portugal: Sociedade Portuguesa de Inovação**, 2006.

SILVA, P. M. d. Modelo de aceitação de tecnologia (tam) aplicado ao sistema de informação da biblioteca virtual em saúde (bvs) nas escolas de medicina da região metropolitana do recife. PPGCI/UFPB, 2008.

SPRAGUE, R. H.; WATSON, H. J. **Sistema de Apoio à Decisão: colocando a teoria em prática**. [S.l.]: Campus, 1991.

TEOREY, T. J. et al. **Database Modeling and Design**. 4. ed. [S.l.]: Morgan Kaufmann, 2005. (The Morgan Kaufmann Series in Data Management Systems).

The PostgreSQL Global Development Group. **PostgreSQL: The world's most advanced open source database**. 2016. Available from Internet: <<https://www.postgresql.org>>.

TURBAN, E. et al. **Business Intelligence: Um enfoque gerencial para a inteligência do negócio**. [S.l.]: Bookman, 2009.

VERCELLIS, S. **Data mining and optimization for decision making**. [S.l.]: John Wiley & Sons Ltd, 2009.

ZELNY, M. Management support systems: towards integrated knowledge management. **Human systems management**, IOS Press, v. 7, n. 1, p. 59–70, 1987.

ZUCCA, J. Business intelligence infrastructure for academic libraries. **Evidence Based Library And Information Practice**, v. 8, n. 2, 2013.

**APÊNDICE A — DOCUMENTO DE REQUISITOS**

# **Documento de Requisitos do Sistema**

## **Versão 1.0**

## Conteúdo

<b>1. INTRODUÇÃO .....</b>	<b>3</b>
1.1 VISÃO GERAL DO DOCUMENTO .....	3
1.2 CONVENÇÕES, TERMOS E ABREVIACÕES .....	3
1.2.1 Identificação dos requisitos .....	3
1.2.2 Prioridades dos requisitos .....	3
<b>2. DESCRIÇÃO GERAL DO SISTEMA .....</b>	<b>4</b>
2.1 STAKEHOLDERS .....	4
<b>3. REQUISITOS FUNCIONAIS .....</b>	<b>4</b>
[RF001] Registros duplicados.....	4
[RF002] Exemplares iguais com classificação diferente .....	4
[RF003] Exemplares nacionais nunca emprestados .....	5
[RF004] Exemplares importados nunca emprestados .....	5
[RF005] Campi com menor número de empréstimos.....	5
[RF006] Campi com maior número de empréstimos .....	5
[RF007] Horários com maior número de empréstimos .....	5
[RF008] Exemplares sem registro do valor de compra .....	5
[RF009] Exemplares por área do CNPq.....	6
<b>4. REQUISITOS NÃO-FUNCIONAIS .....</b>	<b>6</b>
[NF001] Usabilidade .....	6
[NF002] Segurança.....	6
[NF003] Desempenho .....	6
[NF003] Licença .....	6
[NF003] Interface .....	6
<b>5. GERENCIAMENTO DE REQUISITOS .....</b>	<b>7</b>

## 1. Introdução

Este documento especifica os requisitos da solução de *Business Intelligence* desenvolvida para as bibliotecas da Universidade Federal do Pampa, fornecendo aos desenvolvedores as informações necessárias para o projeto e implementação, assim como para a realização dos testes e homologação do sistema.

### 1.1 Visão geral do documento

Além desta seção introdutória, as seções seguintes estão organizadas como descrito abaixo.

**Seção 2 – Descrição geral do sistema:** apresenta uma visão geral do sistema, caracterizando qual é o seu escopo e descrevendo seus *stakeholders*.

**Seção 3 – Requisitos funcionais:** especifica os casos de uso a serem implementados no sistema, descrevendo suas prioridades.

**Seção 4 – Requisitos não-funcionais:** especifica todos os requisitos não funcionais do sistema, divididos em requisitos de usabilidade, interface, desempenho, segurança, e tipo de licença dos *softwares* utilizados.

**Seção 5 – Gerenciamento de requisitos:** apresenta as etapas a serem seguidas para a alteração dos requisitos apresentados no documento.

### 1.2 Convenções, termos e abreviações

A correta interpretação deste documento exige o conhecimento de algumas convenções e termos específicos, que são descritos a seguir.

#### 1.2.1 Identificação dos requisitos

Por convenção, a referência a requisitos é feita através do nome da subseção onde eles estão descritos, seguidos do identificador do requisito, de acordo com a especificação a seguir:

[nome da subseção. identificador do requisito]

Os requisitos devem ser identificados com um identificador único. A numeração inicia com o identificador [RF001] ou [NF001] e prossegue sendo incrementada à medida que forem surgindo novos requisitos.

#### 1.2.2 Prioridades dos requisitos

Para estabelecer a prioridade dos requisitos, nas seções 4 e 5, foram adotadas as denominações “essencial”, “importante” e “desejável”.

**Essencial** é o requisito sem o qual o sistema não entra em funcionamento. Requisitos essenciais são requisitos imprescindíveis, que têm que ser implementados impreterivelmente.

**Importante** é o requisito sem o qual o sistema entra em funcionamento, mas de forma não satisfatória. Requisitos importantes devem ser implementados, mas, se não forem, o sistema poderá ser implantado e usado mesmo assim.

**Desejável** é o requisito que não compromete as funcionalidades básicas do sistema, isto é, o sistema pode funcionar de forma satisfatória sem ele. Requisitos desejáveis podem ser deixados para versões posteriores do sistema, caso não haja tempo hábil para implementá-los na versão que está sendo especificada.

## 2. Descrição geral do sistema

Apesar de existir uma grande quantidade de dados relativos às bibliotecas no banco de dados da Universidade Federal do Pampa (UNIPAMPA), o atual sistema de bibliotecas não é capaz de oferecer informações de forma eficiente para auxiliar no processo decisório dos gestores. Esse problema afeta indiretamente todos os usuários das bibliotecas, uma vez que interfere na qualidade dos serviços prestados por elas e na maneira como são investidos seus recursos.

A solução proposta, consiste no desenvolvimento de um sistema de suporte a tomada de decisão, baseado em técnicas de *Business Intelligence* (BI). Esse sistema deve ser responsável por conectar-se com a base de dados da UNIPAMPA, extrair os dados necessários, processá-los, carregá-los para um *Data Warehouse* (DW) e, com o uso de técnicas de exploração de dados, deve ser capaz de apresentar informações de forma eficiente, de forma a auxiliar o processo decisório dos gestores das bibliotecas da UNIPAMPA.

### 2.1 Stakeholders

Os *stakeholders* do projeto podem ser divididos em dois grupos: os desenvolvedores, responsáveis pelo projeto e desenvolvimento da solução e os usuários, que utilizarão o sistema desenvolvido. Os componentes de cada grupo são:

- Usuários: Gestores das bibliotecas da UNIPAMPA;
- Desenvolvedores: Projetista e orientador.

## 3. Requisitos funcionais

### [RF001] Registros duplicados

**Descrição:** A solução deve apresentar informação sobre quais são os exemplares com mesma edição e ISBN que apareçam em registros diferentes.

**Prioridade:**     Essencial                       Importante                       Desejável

### [RF002] Exemplares iguais com classificação diferente

**Descrição:** A solução deve apresentar informação sobre quais são os exemplares com mesmo ISBN e edição que possuem classificação diferente.

**Prioridade:**     Essencial                       Importante                       Desejável

### [RF003] Exemplos nacionais nunca emprestados

**Descrição:** A solução deve apresentar informação sobre quantos e quais são os exemplares nacionais nunca emprestados em cada *campus*.

**Prioridade:**  Essencial  Importante  Desejável

### [RF004] Exemplos importados nunca emprestados

**Descrição:** A solução deve apresentar informação sobre quantos e quais são os livros importados nunca emprestados em cada *campus*.

**Prioridade:**  Essencial  Importante  Desejável

### [RF005] Campi com menor número de empréstimos

**Descrição:** A solução deve apresentar informação sobre quais os *campi* com menor número de empréstimos.

**Prioridade:**  Essencial  Importante  Desejável

### [RF006] Campi com maior número de empréstimos

**Descrição:** A solução deve apresentar informação sobre quais os *campi* com maior número de empréstimos.

**Prioridade:**  Essencial  Importante  Desejável

### [RF007] Horários com maior número de empréstimos

**Descrição:** A solução deve apresentar informação sobre quais os horários com maior número de empréstimos em cada *campus*.

**Prioridade:**  Essencial  Importante  Desejável

### [RF008] Exemplos sem registro do valor de compra

**Descrição:** A solução deve apresentar informação sobre quais são os exemplares que estão sem registro de valor de compra em cada *campus*.

**Prioridade:**  Essencial  Importante  Desejável

### [RF009] Exemplos por área do CNPq

**Descrição:** A solução deve apresentar informação sobre quantos exemplares existem de cada área do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

**Prioridade:**  Essencial  Importante  Desejável

## 4. Requisitos não-funcionais

### [NF001] Usabilidade

A aplicação deve ter uma interface amigável e ser de fácil manuseio para os usuários.

**Prioridade:**  Essencial  Importante  Desejável

### [NF002] Segurança

A solução deve possibilitar acesso às informações, através de login e senha, apenas para usuários pré cadastrados.

**Prioridade:**  Essencial  Importante  Desejável

### [NF003] Desempenho

O tempo de resposta da solução não deve ser superior a 5 segundos.

**Prioridade:**  Essencial  Importante  Desejável

### [NF003] Licença

Todo código fonte e ferramentas utilizadas para compor a solução devem ser *open source*.

**Prioridade:**  Essencial  Importante  Desejável

### [NF003] Interface

A solução deve ter uma interface de apresentação web, para que vários usuários possam acessá-la simultaneamente através de um navegador web.

**Prioridade:**  Essencial  Importante  Desejável

## **5. Gerenciamento de Requisitos**

Para alteração ou inclusão de requisitos no projeto da solução deverão ser seguidas as etapas descritas a seguir:

- O usuário solicita uma mudança de requisito ao projetista;
- O projetista recebe a mudança sugerida;
- O projetista juntamente com o orientador do projeto analisarão tal mudança e avaliarão o impacto da mesma no sistema;
- O projetista juntamente com o cliente negociam a mudança pretendida;
- Como resultado dessa negociação ocorrerá ou não a mudança solicitada.

**APÊNDICE B — CARTA DE CIÊNCIA**

**CARTA DE CIÊNCIA SOBRE TRABALHO DE CONCLUSÃO DE CURSO E  
ACEITE QUANTO A COLETA DE DADOS NA UNIVERSIDADE**

Na condição de Reitor da Universidade Federal do Pampa (CPF 209.809.570-87), tenho conhecimento da proposta de Trabalho de Conclusão de Curso **“Aplicando técnicas de business intelligence sobre dados das bibliotecas da Unipampa”**, a ser desenvolvido pelo Acadêmico do Curso de Engenharia da Computação, Leonardo Gauer Schulte, sob a Coordenação do Prof. Sandro da Silva Camargo.

Nesse sentido estamos de acordo com sua operacionalização e com a coleta de dados nesta Universidade. Autorizo o desenvolvimento da pesquisa, para fins de consolidação do trabalho proposto.

Bagé, 07 de julho de 2016.



Marco Antonio Fontoura Hansen  
Reitor

## APÊNDICE C — FUNÇÕES DE TRATAMENTO E CARGA DOS DADOS

```

CREATE TRIGGER gatilhoBib
AFTER INSERT ON intermediario.biblioteca
FOR EACH ROW EXECUTE PROCEDURE intermediario.sincronizabib ( );

CREATE OR REPLACE FUNCTION intermediario.sincronizaBib ( )
RETURNS TRIGGER AS $$ DECLARE
BEGIN
    INSERT INTO public. "dimBiblioteca" (
        id,
        nome )
    VALUES (
        new.id,
        new.nome );
RETURN NULL;
END;
$$ LANGUAGE 'plpgsql';

```

---

```

CREATE TRIGGER gatilhoReg
AFTER INSERT ON intermediario.registro
FOR EACH ROW EXECUTE PROCEDURE intermediario.sincronizaReg ( );

CREATE OR REPLACE FUNCTION intermediario.sincronizaReg ( )
RETURNS TRIGGER AS $$ DECLARE a INT;
e INT;
BEGIN
BEGIN
    a = (
        regexp_matches (
            new. "anoBibliografico",
            '\d+' ) )
    [ 1 ] ::INT;
EXCEPTION

```

```

WHEN OTHERS THEN a = NULL;
END;
BEGIN
    e = (
        regexp_matches (
            new.edicao,
            '\d+' ) )
    [ 1 ] ::INT;
EXCEPTION
WHEN OTHERS THEN e = NULL;
END;
INSERT INTO public. "dimTitulo" VALUES (
    new.id,
    NULL,
    new. "numeroControle" ,
    new.titulo,
    new.subtitulo,
    new.autor,
    new. "numeroChamada" ,
    new.local ,
    new.editora,
    a,
    e );
RETURN NULL;
END;
$$ LANGUAGE 'plpgsql';

```

---

```

CREATE TRIGGER gatilhoItm
AFTER INSERT ON intermediario.item
FOR EACH ROW EXECUTE PROCEDURE intermediario.sincronizaItm ( );

CREATE OR REPLACE FUNCTION intermediario.sincronizaItm ( )
RETURNS TRIGGER AS $$ DECLARE c1 INT;
BEGIN
    IF REPLACE (

```

```

        lower (
            new. "consultaLocal" ),
        ' ',
        '' )
    = 'consultalocal' THEN cl = 1;
ELSE cl = 0;
END IF;
INSERT INTO public. "dimfatoExemplar" VALUES (
    new.id,
    new. "idRegistro",
    new. "idBiblioteca",
    new. "codigoBarras",
    new.classificacao,
    new.cutter,
    cl,
    new. "dtCompra",
    new. "vlCompra",
    0 );
RETURN NULL;
END;
$$ LANGUAGE 'plpgsql';

```

---

```

CREATE TRIGGER gatilhoEmp
AFTER INSERT ON intermediario.emprestimo
FOR EACH ROW EXECUTE PROCEDURE intermediario.sincronizaEmp ( );

CREATE OR REPLACE FUNCTION intermediario.sincronizaEmp ( )
RETURNS TRIGGER AS $$ DECLARE dr INT;
hr INT;
dd INT;
hd INT;
aux INT;
m INT;
h INT;
BEGIN

```

*--Tabela dimFuncionario*

```

IF NOT EXISTS (
    SELECT
        id
    FROM
        public. "dimFuncionario"
    WHERE
        id = new. "idBibliotecarioRet" )
THEN
    IF new. "idBibliotecarioRet" IS NOT NULL
    THEN INSERT INTO public. "dimFuncionario" VALUES (
        new. "idBibliotecarioRet",
    NULL,
    NULL );
```

**END IF;**

**END IF;**

```

IF NOT EXISTS (
    SELECT
        id
    FROM
        public. "dimFuncionario"
    WHERE
        id = new. "idBibliotecarioDev" )
THEN
    IF new. "idBibliotecarioDev" IS NOT NULL
    THEN INSERT INTO public. "dimFuncionario" VALUES (
        new. "idBibliotecarioDev",
    NULL,
    NULL );
```

**END IF;**

**END IF;**

*--Tabela dimUsuario*

```

IF NOT EXISTS (
    SELECT
        id
    FROM
```

```

        public. "dimUsuario"
    WHERE
        id = new. "idUsuario" )
THEN
IF new. "idUsuario" IS NOT NULL
THEN INSERT INTO public. "dimUsuario" VALUES (
    new. "idUsuario",
    NULL,
    NULL,
    NULL );
END IF;
END IF;
—Tabela dimData
aux = (
    SELECT
        id
    FROM
        public. "dimData"
    WHERE
        "data" = new. "dataRetirada" );
IF aux IS NULL THEN INSERT INTO public. "dimData" VALUES (
    DEFAULT,
    new. "dataRetirada",
    extract (
        DAY
    FROM
        new. "dataRetirada" ),
    extract (
        MONTH
    FROM
        new. "dataRetirada" ),
    extract (
        YEAR
    FROM
        new. "dataRetirada" ) )
RETURNING id INTO dr;

```

```

ELSE dr = aux;
END IF;
aux = (
  SELECT
    id
  FROM
    public. "dimData"
  WHERE
    "data" = new. "dataDevolucao" );
IF aux IS NULL THEN INSERT INTO public. "dimData" VALUES (
  DEFAULT,
  new. "dataDevolucao",
  extract (
    DAY
  FROM
    new. "dataDevolucao" ),
  extract (
    MONTH
  FROM
    new. "dataDevolucao" ),
  extract (
    YEAR
  FROM
    new. "dataDevolucao" ) )
RETURNING id INTO dd;
ELSE dd = aux;
END IF;
—Tabela dimHora
—Arredondamento
m = extract (
  MINUTE
  FROM
  new. "horaRetirada" );
h = extract (
  HOUR
  FROM

```

```

        new. "horaRetirada" );
IF m < 7 THEN m = 0;
ELSE IF m < 22 THEN m = 15;
ELSE IF m < 37 THEN m = 30;
ELSE IF m < 52 THEN m = 45;
ELSE m = 0;
h = h + 1;
END IF;
END IF;
END IF;
END IF;
—Escrever na tabela
aux = (
    SELECT
        id
    FROM
        public. "dimHora"
    WHERE
        "horario" = to_timestamp (
            (
                h || ':' || m || ':' || '00' ),
            'hh24:mi:ss' ) ::time WITHOUT time ZONE );
IF aux IS NULL THEN INSERT INTO public. "dimHora" VALUES (
    DEFAULT,
    to_timestamp (
        (
            h || ':' || m || ':' || '00' ),
            'hh24:mi:ss' ) ::time WITHOUT time ZONE,
    h,
    m )
RETURNING id INTO hr;
ELSE hr = aux;
END IF;
—Arredondamento
m = extract (
    MINUTE

```

```

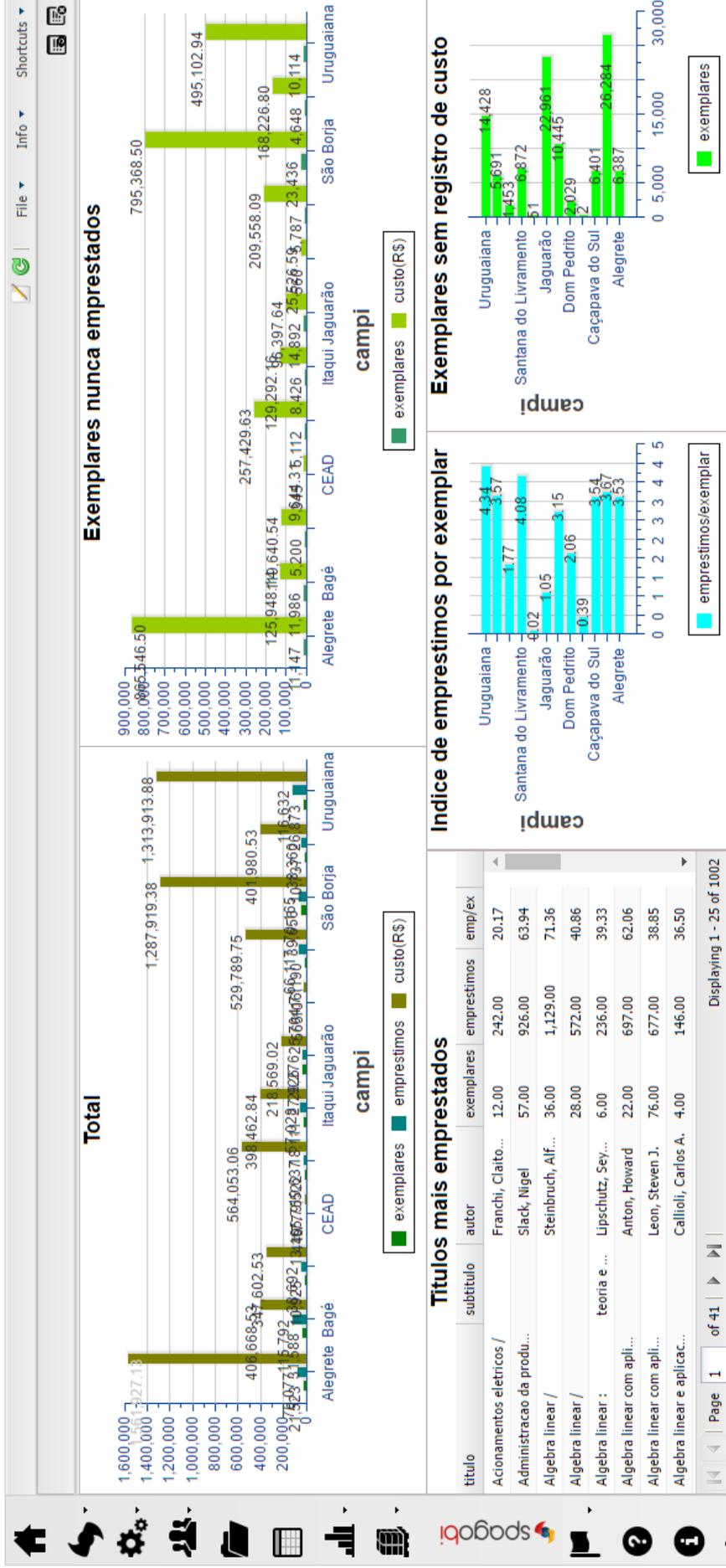
    FROM
        new. "horaDevolucao" );
h = extract (
    HOUR
    FROM
        new. "horaDevolucao" );
IF m < 7 THEN m = 0;
ELSE IF m < 22 THEN m = 15;
ELSE IF m < 37 THEN m = 30;
ELSE IF m < 52 THEN m = 45;
ELSE m = 0;
h = h + 1;
END IF;
END IF;
END IF;
END IF;
—Escrever na tabela
aux = (
    SELECT
        id
    FROM
        public. "dimHora"
    WHERE
        "horario" = to_timestamp (
            (
                h || ':' || m || ':' || '00' ),
            'hh24:mi:ss' ) ::time WITHOUT time ZONE );
IF aux IS NULL THEN INSERT INTO public. "dimHora" VALUES (
    DEFAULT,
    to_timestamp (
        (
            h || ':' || m || ':' || '00' ),
            'hh24:mi:ss' ) ::time WITHOUT time ZONE,
    h,
    m )
RETURNING id INTO hd;

```

```
ELSE hd = aux;
END IF;
—Tabela fatoEmprestimo
INSERT INTO public. "fatoEmprestimo" VALUES (
    new. "id",
    dr,
    dd,
    hr,
    hd,
    new. "idUsuario",
    new. "idBibliotecarioRet",
    new. "idBibliotecarioDev",
    new. "idBiblioteca",
    new. "idItem" );
UPDATE public. "dimfatoExemplar"
SET
    emprestado = emprestado+1
WHERE
    id = new. "idItem";
RETURN NULL;
END;
$$ LANGUAGE 'plpgsql';
```

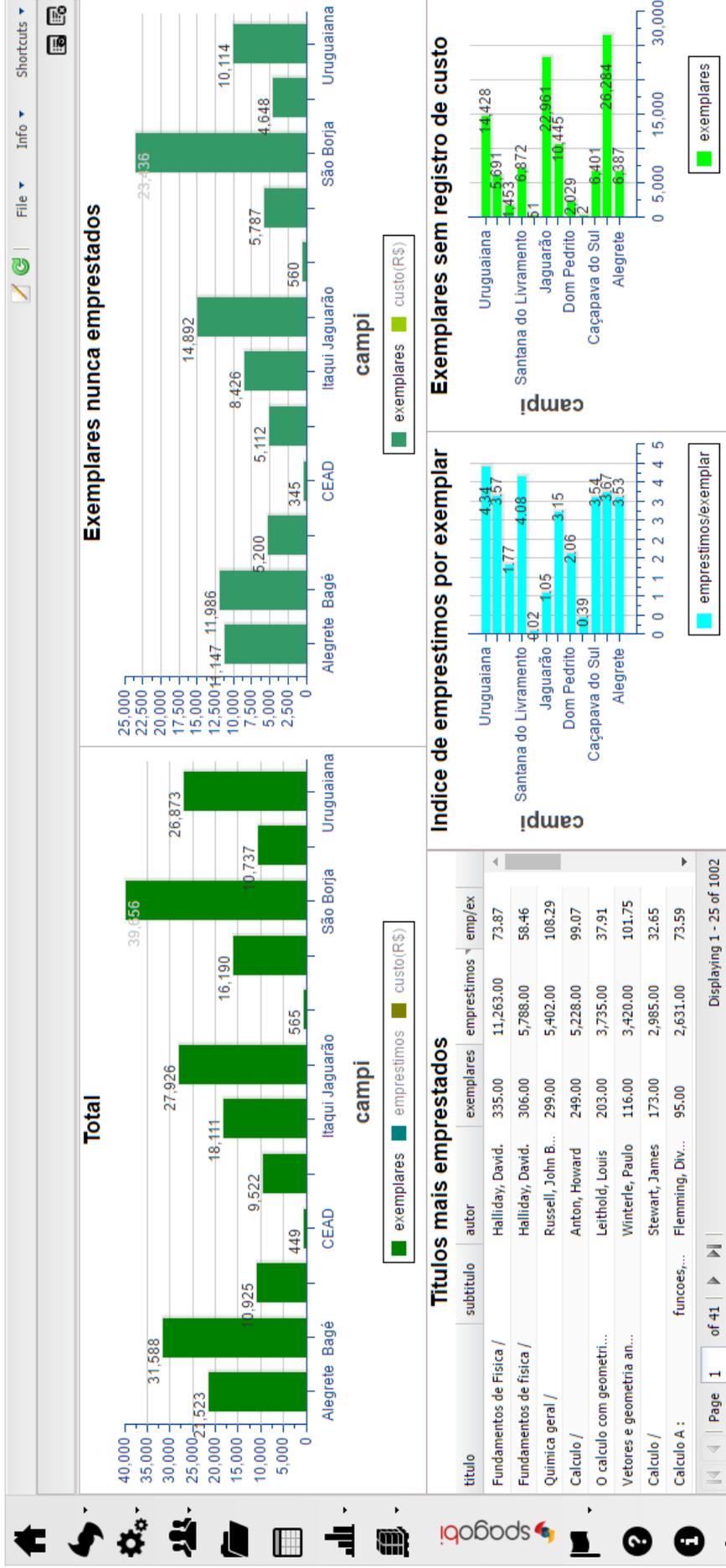
**APÊNDICE D — INTERFACES DA SOLUÇÃO**

Figura 1 – Dashboard de acervo completo



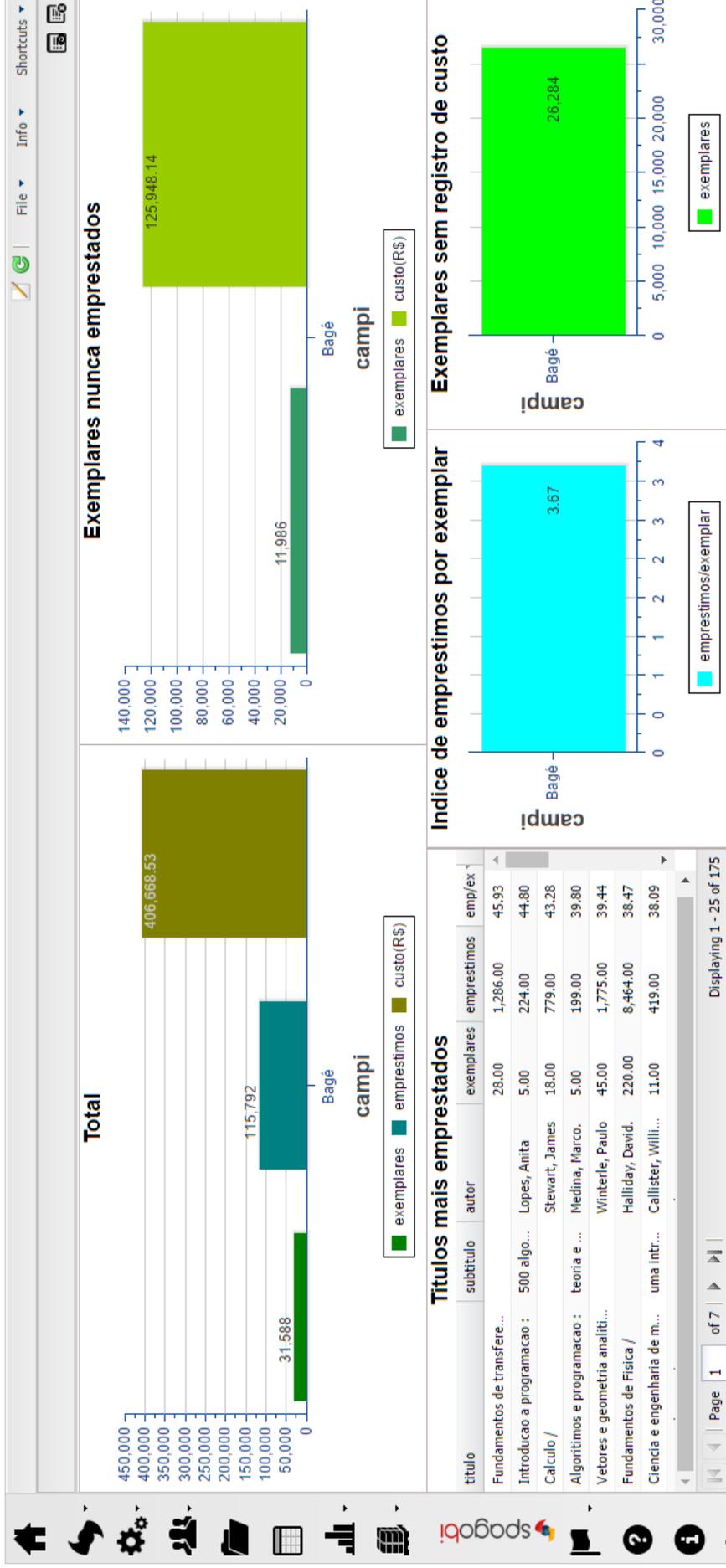
Fonte: Próprio Autor (2016)

Figura 2 – Dashboard de acervo mostrando a quantidade de exemplares



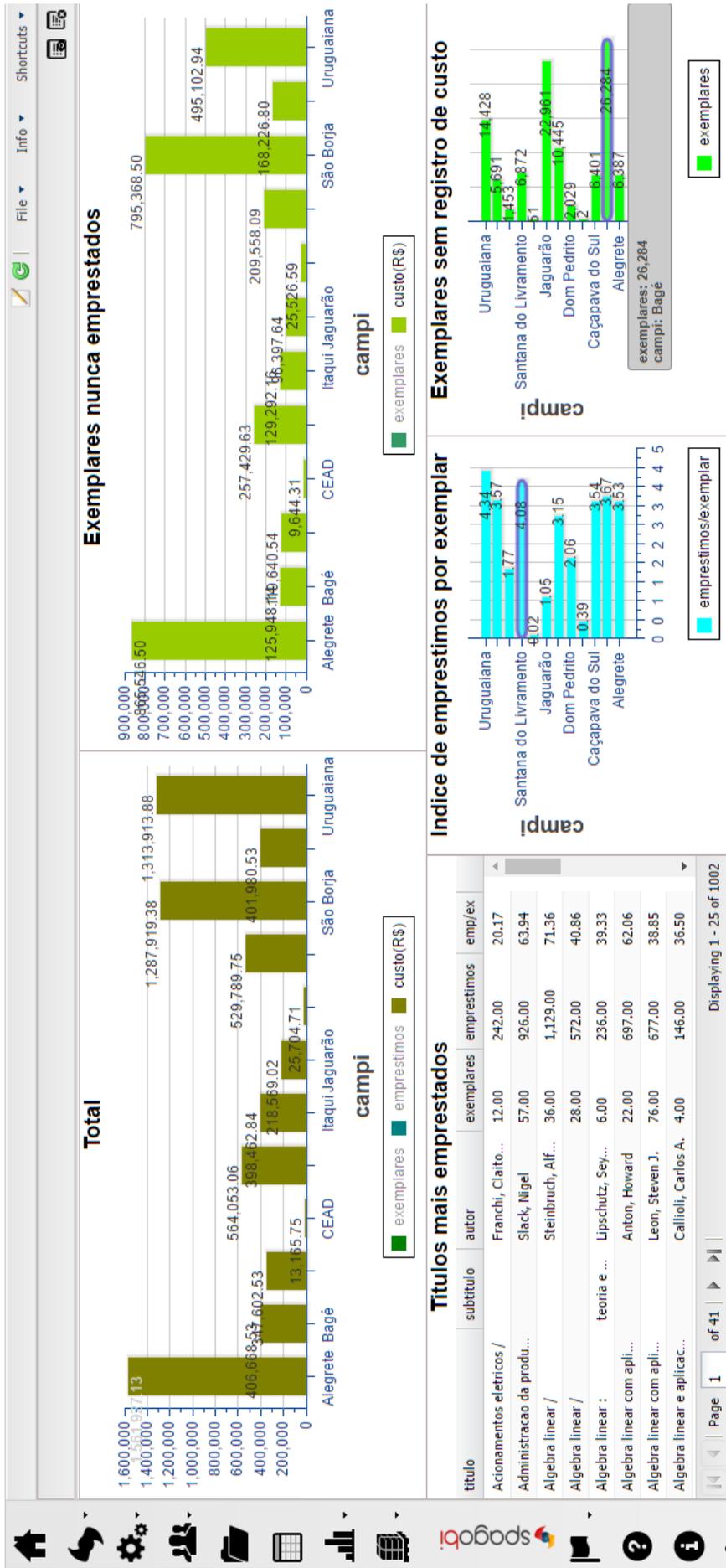
Fonte: Próprio Autor (2016)

Figura 3 – Dashboard de acervo filtrado para o *campus* Bagé



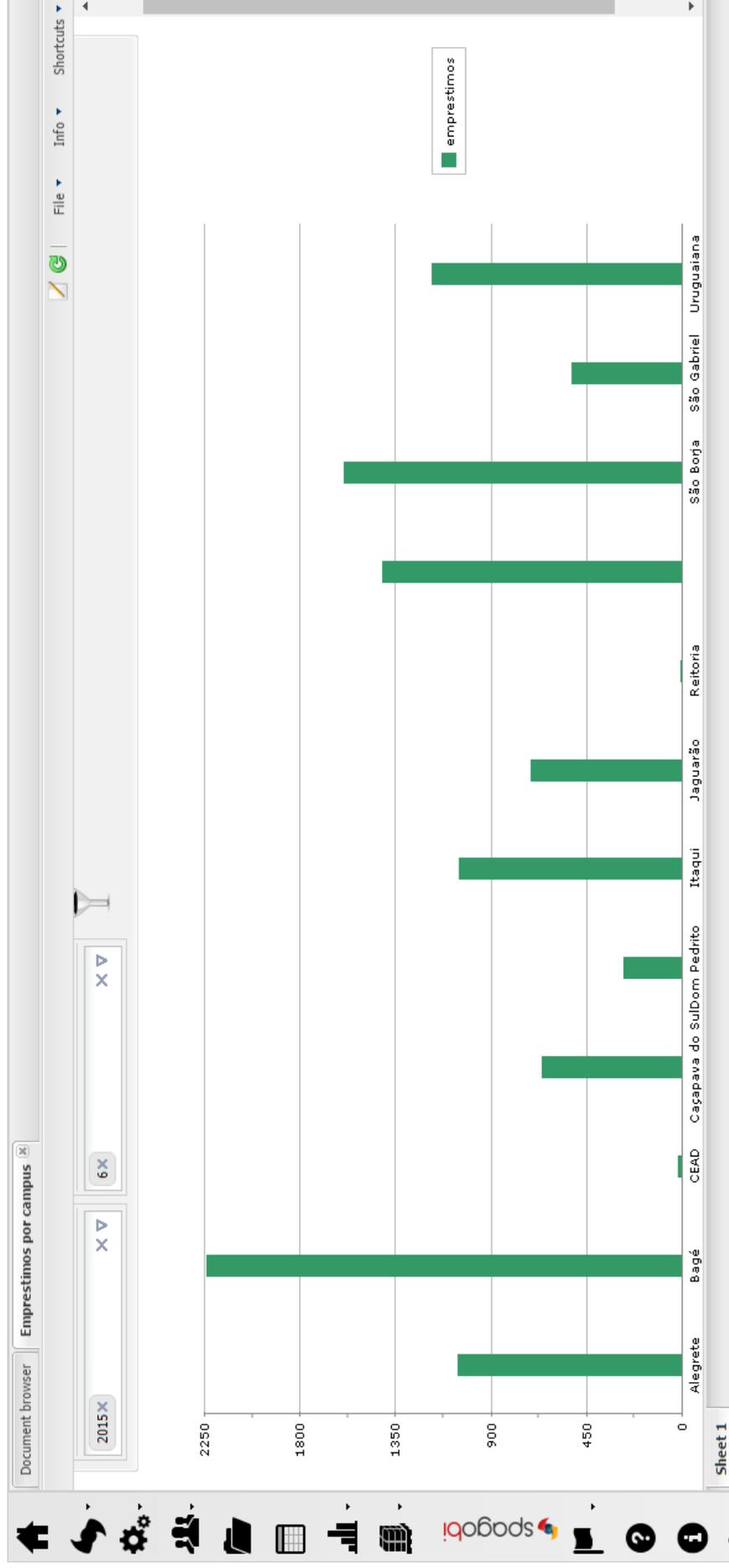
Fonte: Próprio Autor (2016)

Figura 4 – Dashboard de acervo mostrando o custo dos exemplares



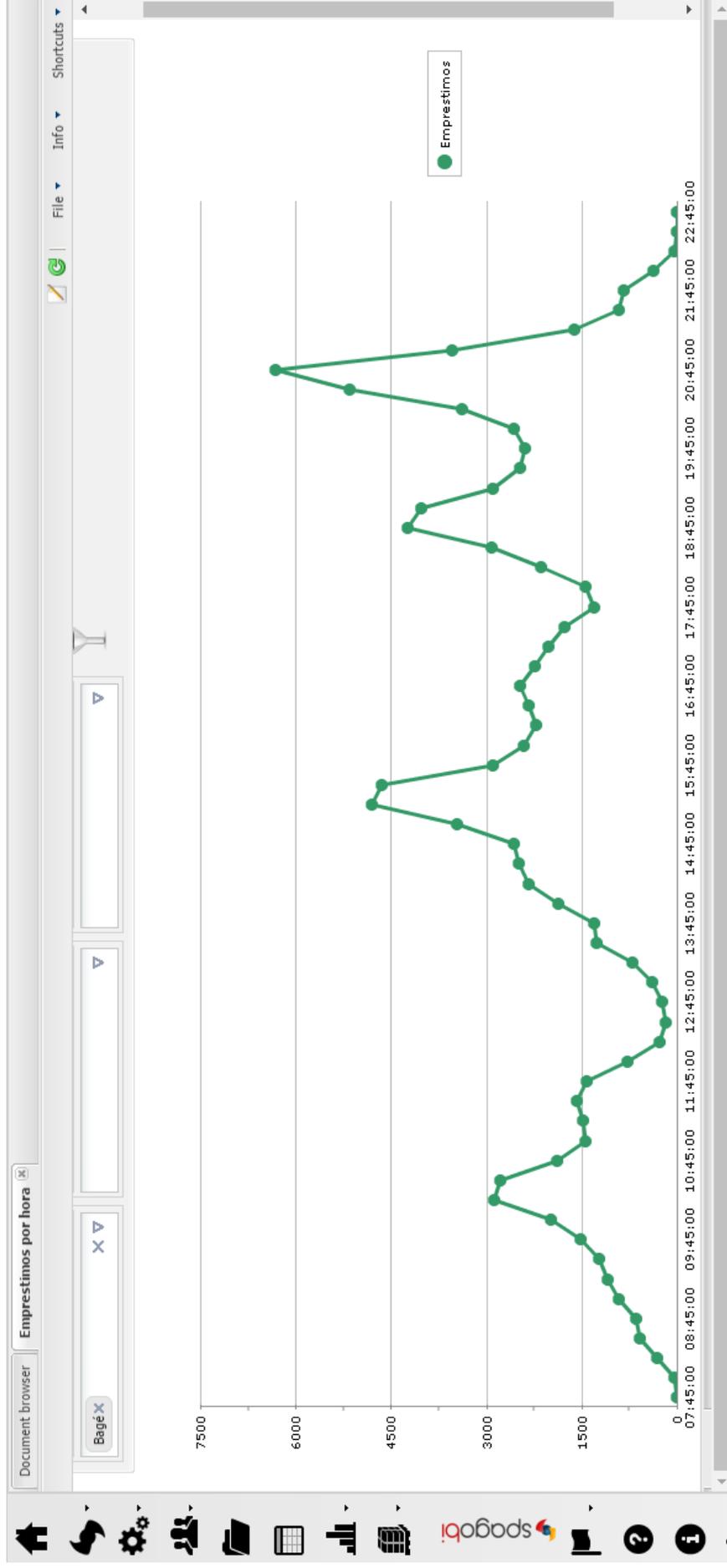
Fonte: Próprio Autor (2016)

Figura 5 – Empréstimos por campus no ano de 2015



Fonte: Próprio Autor (2016)

Figura 6 – Empréstimos por horário no *campus* Bagé



Fonte: Próprio Autor (2016)

Figura 7 – Relatório de exemplares do *campus* Bagé nunca emprestados

Document browser Exemplares nunca emprestados

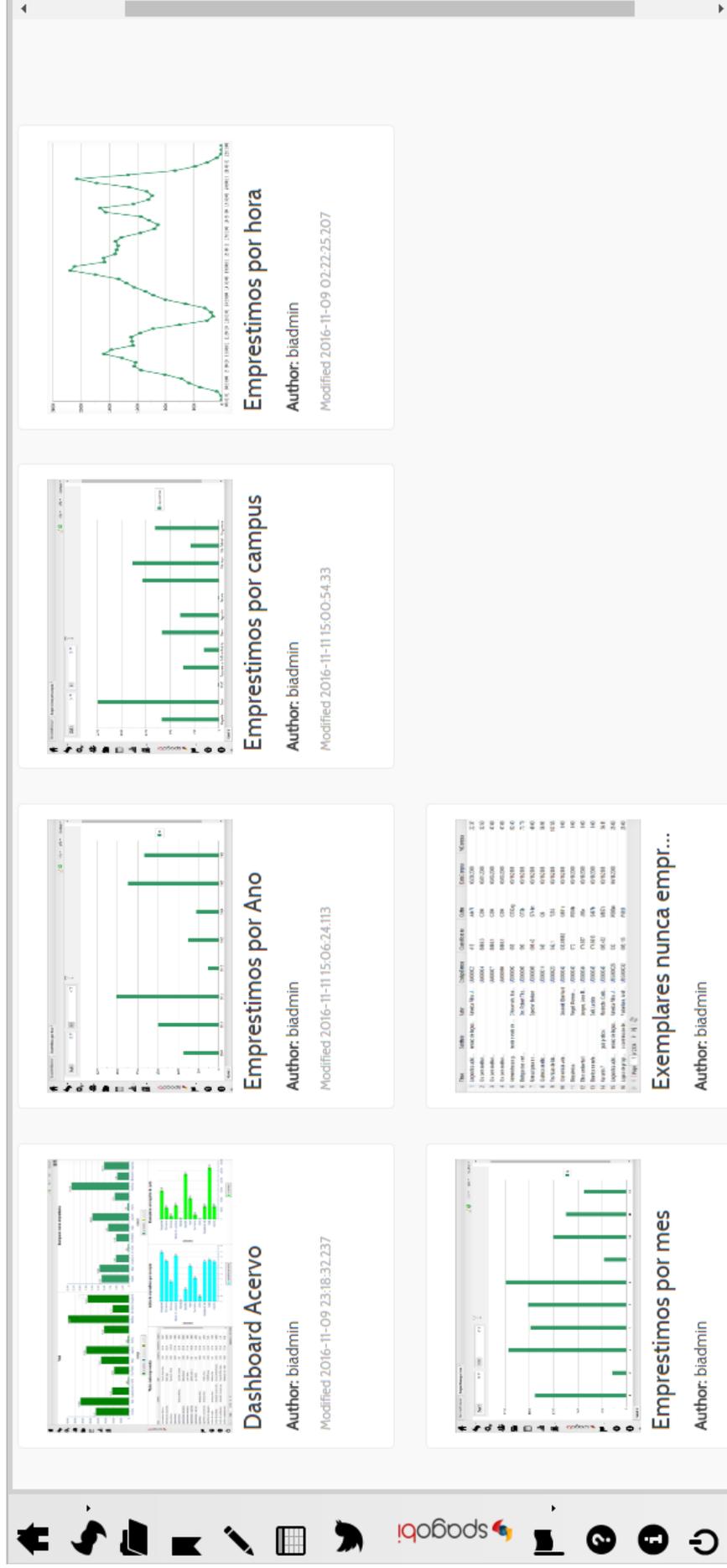
campi Bagé X

	Título	Subtítulo	Autor	CodigoBarras	Classificacao	Cutter	DataCompra	VCCompra
1	Linguistica aplic...	ensino de lingua...	Almeida Filho, J...	UBG000026	530	R585m	04/18/2008	25,00
2	Logica de progr...	a construo de...	Forbellone, Andr...	UBG000032	005.115	F692I		25,00
3	Logica de progr...	a construo de...	Forbellone, Andr...	UBG000033	005.115	F692I		25,00
4	A aventura do li...	do leitor ao nave...	Chartier, Roger,	UBG000045	002	C486a	03/10/2008	53,00
5	A aventura do li...	do leitor ao nave...	Chartier, Roger,	UBG000046	002	C486a	03/10/2008	53,00
6	A aventura do li...	do leitor ao nave...	Chartier, Roger,	UBG000047	002	C486a	03/10/2008	53,00
7	A aventura do li...	do leitor ao nave...	Chartier, Roger,	UBG000048	002	C486a	03/10/2008	53,00
8	A aventura do li...	do leitor ao nave...	Chartier, Roger,	UBG000049	002	C486a	03/10/2008	53,00
9	A aventura do li...	do leitor ao nave...	Chartier, Roger,	UBG000050	002	C486a	03/10/2008	53,00
10	A aventura do li...	do leitor ao nave...	Chartier, Roger,	UBG000051	002	C486a	03/10/2008	53,00
11	A aventura do li...	do leitor ao nave...	Chartier, Roger,	UBG000052	002	C486a	03/10/2008	53,00
12	Quimica Organi...		Bruice, Paula Yu...	UBG000068	547	B886q	03/10/2008	100,00
13	Quimica Organi...		Bruice, Paula Yu...	UBG000069	547	B886q	03/10/2008	100,00
14	Boas Maneiras ...	um guia pratico ...	Ribeiro, Celia.	UBG000079	395.52	R484b	03/10/2008	30,00
15	Os persas /		Esquilo	UBG000206	880	E77p		37,00
16	Os persas /		Esquilo	UBG000207	880	E77p		37,00

Page 1 of 273 Sheet 1

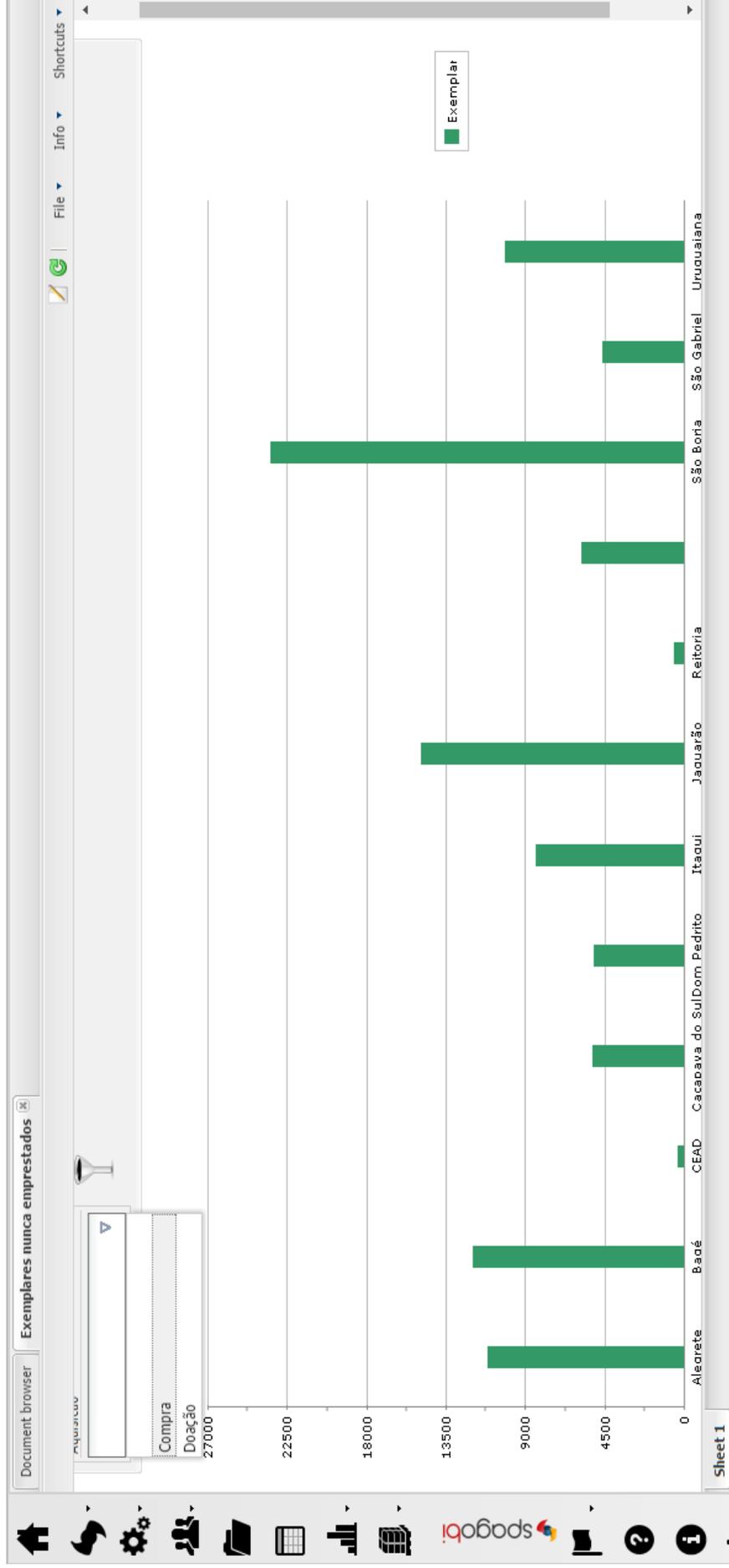
Fonte: Próprio Autor (2016)

Figura 8 – Área de análise



Fonte: Próprio Autor (2016)

Figura 9 – Exemplares nunca emprestados por campus



Fonte: Próprio Autor (2016)

## APÊNDICE E — DOCUMENTO DE AVALIAÇÃO DA SOLUÇÃO

### Questionário de aceitação de tecnologia

1- Foi fácil aprender a usar a ferramenta de *Business Intelligence*.

- Concordo totalmente
- Concordo parcialmente
- Indiferente
- Discordo parcialmente
- Discordo totalmente

2- Considero que a ferramenta de *Business Intelligence* seja de fácil manuseio.

- Concordo totalmente
- Concordo parcialmente
- Indiferente
- Discordo parcialmente
- Discordo totalmente

3- A ferramenta apresenta tempo de resposta satisfatório (tempo necessário para carregar as informações e realizar os filtros).

- Concordo totalmente
- Concordo parcialmente
- Indiferente
- Discordo parcialmente
- Discordo totalmente

4- A ferramenta não apresenta lentidão ou travamentos.

- Concordo totalmente
- Concordo parcialmente
- Indiferente
- Discordo parcialmente
- Discordo totalmente

5- A interface da ferramenta facilita a compreensão das informações apresentadas.

- Concordo totalmente

Concordo parcialmente

Indiferente

Discordo parcialmente

Discordo totalmente

6- A solução de *Business Intelligence* apresenta informações relevantes.

Concordo totalmente

Concordo parcialmente

Indiferente

Discordo parcialmente

Discordo totalmente

7- A solução de *Business Intelligence* auxilia no processo decisório.

Concordo totalmente

Concordo parcialmente

Indiferente

Discordo parcialmente

Discordo totalmente

8- Pretendo utilizar a ferramenta diariamente.

Concordo totalmente

Concordo parcialmente

Indiferente

Discordo parcialmente

Discordo totalmente

9- Pretendo realizar ações para melhorar o emprego de recursos e os serviços prestados nas bibliotecas da UNIPAMPA, com base no conhecimento descoberto pela solução de *Business Intelligence*.

Concordo totalmente

Concordo parcialmente

Indiferente

Discordo parcialmente

Discordo totalmente

