

Aplicando Técnicas de Business Intelligence sobre dados de desempenho Acadêmico: Um estudo de caso

Ana Magela Rodriguez Almeida¹, Sandro da Silva Camargo¹

¹ Curso Engenharia de Computação– Universidade Federal do Pampa (UNIPAMPA)
Caixa Postal 07 – 96.400-970 – Bagé – RS – Brasil

{anamagelaa, camargo.sandro}@gmail.com

Abstract. *In the present days, the volumes of data generated and stored by organizations are growing disproportionately. This situation comes along with big challenges, like the integration and transformation of this data in relevant information to optimize the process of decision-making. So it's necessary to build a knowledge base to integrate and organize the available data in order to make them easy to understand and allow the quick access, enabling exploitation and analysis of data in order to meet the strategic expectations of organizations. In this scenario, this paper presenting an approach for implement Business Intelligence in the UNIPAMPA, aiming to significantly improve the resource management application in the organization.*

Resumo. *Atualmente, os volumes de dados gerados e armazenados pelas organizações estão crescendo de forma desmesurada. Esta situação, vem acompanhada de grandes desafios, entre eles, a integração e transformação destes dados em informações relevantes para aprimoramento do processo decisório. Desta forma, faz-se necessária a execução de um complexo processo para organizar e integrar os dados disponíveis de forma facilmente entendida e que permita o acesso rápido, possibilitando a exploração e análise de dados a fim de atender as expectativas estratégicas das organizações. Neste cenário, este trabalho apresenta uma abordagem para implementação de Business Intelligence dentro de uma instituição federal de ensino superior, a fim de otimizar a aplicação de seus recursos.*

1. Introdução

O sucesso de uma organização é dependente de sua capacidade tomar decisões corretas. Assim, o processo de tomada de decisão é um aspecto chave dentro da gestão organizacional. A existência de dados e informações que descrevam o contexto do problema auxilia a maximizar o sucesso das decisões tomadas. Paralelamente, ao longo de sua existência, as organizações geram e armazenam uma grande quantidade de dados referentes as atividades realizadas. Desta forma, a aplicação de técnicas de business intelligence (BI) no processo de tomada de decisão estratégica torna-se de extrema importância, uma vez que decisões equivocadas podem comprometer o futuro de uma organização.

Inseridas no mesmo contexto das demais organizações, instituições de Ensino também necessitam ter recursos para subsidiarem a otimização de seus processos decisórios. Apesar de ter apenas 8 anos, a Universidade Federal do Pampa (UNIPAMPA) já armazena em seu banco de dados mais de um milhão de registros que contam sua história, apontando seus problemas e virtudes. Porém, é necessário

transformar estes dados armazenados em informações úteis para a Instituição. Desta forma, a implementação de soluções BI aplicadas em organizações de ensino podem ser de alta relevância para identificar, compreender e prever possíveis problemas residentes na Instituição, que muitas vezes são resultado de decisões tomadas sem base no conhecimento escondido nestes dados. Desta maneira, a consolidação e exploração destes dados apoiam o processo de tomada de decisões e assim mostram novas estratégias a fim de otimizar a utilização dos recursos da Instituição.

Com base na complexidade em gerenciar e processar grande volume de dados, este trabalho está focado na aplicabilidade de técnicas no contexto BI sobre os dados de desempenho acadêmico de uma instituição de ensino superior. Desta forma, este trabalho visa descrever uma metodologia de aplicação de técnicas de BI com o objetivo de fornecer subsídios para o trabalho dos gestores educacionais, contribuindo para o aprimoramento do processo decisório em uma instituição de ensino superior.

O artigo está estruturado da seguinte forma, na seção 2 é apresentada uma introdução sobre conceitos relevantes a BI, na seção seguinte é descrito o processo de implementação de BI, juntamente com as etapas de criação de *Data Marts* (DM). Na seção 4 é descrito o estudo de caso realizado a partir da plataforma selecionada, assim como uma discussão sobre os resultados.

2. Conceito de Business Intelligence

Business Intelligence consiste em um conjunto de tecnologias, técnicas, conceitos e ferramentas orientadas para análise e apresentação de informações para auxiliar os gestores no processo decisório e como isto permitir às organizações otimizar seus recursos de negócio e alcançar melhores resultados [Wu et. al. 2007].

O principal objetivo de BI é oferecer acesso aos dados de forma simples e, assim proporcionar aos gestores a capacidade de realizar análises convenientes [Turban et. al., 2009]. Ainda para o autor, o processo de BI fundamenta-se na transformação de dados em informações, depois em decisões e, finalmente, em ações. Assim, o propósito de BI é converter volumes de dados em informações novas e úteis, que transformadas em conhecimento podem beneficiar às atividades de uma organização. Para realizar o processo de transformação dos dados em informações, é necessário um ambiente que consolide e permita o uso estratégico dos dados extraídos das bases de dados. Desta maneira, estes ambientes armazenam os dados possibilitando o seu processamento por ferramentas especiais.

2.1. Sistemas de BI

No mercado globalizado existente nos dias de hoje, as organizações cada vez mais buscam destacarem-se e tornarem-se mais competitivas no mercado. Em consequência disto, conseguem otimizar seu desempenho e obterem melhores resultados.

Neste cenário, sistemas de BI proporcionam um ambiente para a unificação dos dados e execução do processo de descoberta de conhecimento. São capazes de extrair, armazenar, processar e interpretar os dados, muitas vezes em tempo real. Estes sistemas podem ser definidos como ferramentas através das quais é possível descobrir conhecimento sobre o histórico operacional da organização, a fim de proporcionar subsídios para uma tomada de decisão mais efetiva pelos gestores e, com isso, tornar este processo mais preciso e confiável.

Segundo [Turban et. al. 2009] e [Cano 2007] Sistemas de BI possuem quatro grandes componentes:

- Fontes de Informação: São sistemas dos quais são obtidas as informações, tais como, sistemas operacionais e transacionais da organização.
- Processo de extração, transformação e carga (ETL): Recupera e transforma os dados que serão carregados para a base que consolida os dados para análise.
- *Data Warehouse* (DW): Repositório de dados integrados e não volátil que proporciona informação preparada no processo ETL para a análise.

Data Mart (DM): É uma base de dados específica para uma determinada área dentro de uma organização.

- Área de apresentação: Conjunto de ferramentas de BI que permitem a exploração e visualização da informação armazenada no DW [Kimball e Ross 2002].

2.2. Data Mart

Um *Data Mart* pode ser definido como uma base de dados departamental específica para uma determinada área dentro de uma organização enfatizando o fácil acesso a uma informação relevante. Para [Kimball e Ross 2002] *Data Mart* são subconjuntos de um *Data Warehouse* completo, possuindo as mesmas características.

Para [Diaz e Caralt 2012], o objetivo de construir um DM é responder a uma determinada análise dentro da organização. Geralmente estes sistemas armazenam menos informação que os DW e permitem acesso rápido das informações para análise, pois possuem indexação de armazenamento. Podem ser tanto dependentes como independentes do DW, constituídos pelas arquiteturas *Top-Down* e *Bottom-Up*.

2.3. Modelagem de Dados

A modelagem é um sistema para concepção e visualização de um modelo de dados. Para a modelagem dos dados armazenados em um *Data warehouse* e em *Data Mart* é utilizada a modelagem dimensional. Para [Kimball e Ross 2002] modelagem dimensional é uma técnica que possibilita a criação de um modelo de dados dimensional.

Este modelo dimensional é constituído por um conjunto de medidas que descrevem aspectos de negócios. Esta modelagem permite sumarizar e estruturar os dados para dar suporte à análise de dados. Três elementos formam este modelo, são eles: Fatos, Dimensões, Medidas.

De acordo com [Kimball e Ross 2002], a tabela de fatos é a principal tabela do modelo. Os fatos são coleções de itens de dados, compostas de dados de medida e de contexto. Estas coleções são compostas pelas medições numéricas que representam a evolução dos negócios de uma organização. O fato registra o dado que será analisado e é composto pela chave primária e um conjunto único de valores de chaves de dimensões.

As tabelas de dimensão contém as descrições de negócio, são os elementos dos fatos do negócio. Cada dimensão pode ter vários níveis hierárquicos para proporcionar um melhor entendimento e uma melhor visualização dos indicadores. Os atributos das dimensões são os principais atributos usados para obter vistas do processo de negócio,

tais como filtro nas consultas, agrupamentos e relatórios [Diaz e Caralt 2012]. As medidas ou variáveis são os atributos numéricos que representam os indicadores que mostram a evolução do negócio da empresa [Turban et. al. 2009].

Existem principalmente duas abordagens dentro da modelagem multidimensional, o modelo *Star Schema* e o modelo *Snow Flake* [Diaz e Caralt 2012]. O *Star Schema* consiste em estruturar informação em processos, vistas e medidas em forma de estrela [Diaz e Caralt 2012]. Tem como característica básica a presença de dados redundantes para proporcionar um melhor desempenho [Junior 2004]. Em termos de desenho, este esquema é composto por uma tabela de fatos no centro para o fato objeto de análise, e uma ou várias tabelas auxiliares chamadas tabelas de dimensões para cada ponto de vista da análise que participa da descrição do fato [Diaz e Caralt 2012]. Neste esquema, a consulta ocorre inicialmente nas tabelas de dimensões e depois na tabela de fatos, garantindo a precisão dos dados através de uma estrutura completa [Junior 2004].

O segundo modelo, o *Snow Flake* é um esquema derivado do modelo *Star Schema*, onde as tabelas de dimensões se normalizam em diversas tabelas. Com isso a tabela de fatos deixa de ser a única tabela que se relaciona com as outras, e assim, surgem novas uniões [Diaz e Caralt 2012].

2.4. Processo de Extração, Transformação e Carga - ETL

O processo ETL é de extrema importância na construção de um sistema BI, pois compreende os procedimentos realizados em torno do DW para coleta e transformação dos dados antes de serem carregados no armazém de dados [Diaz e Caralt 2012]. Este processo consiste na utilização de ferramentas que realizarão estas etapas de coleta, limpeza e migração dos dados ao DW. Segundo [Junior 2004], este processo apresenta-se em três camadas: Extração, Transformação e Carga.

2.5. Ferramentas OLAP

Ferramentas OLAP consistem em um conjunto de técnicas voltadas para acesso e análise *ad-hoc* de dados, utilizando uma serie de recursos para exploração destas informações. Estas aplicações baseadas em *On-Line Analytical Processing*, OLAP, referem-se ao conjunto de processos para integração, análise e manipulação de grande volumes de dados, disponibilizando uma serie de funcionalidades, objetivando uma maior compreensão destes dados por analistas e gestores no processo de análise corporativa.

Os autores [Diaz e Caralt 2012] explicam o conceito como uma tecnologia que permite instanciar os dados em uma visão multidimensional, permitindo a apresentação das informações em distintas perspectivas. A multidimensionalidade é um conceito chave de uma ferramenta OLAP para sintetizar informações, refere-se à visão conceitual personalizada da informação alvo de análise, ou seja, é possível obter diferentes análises a partir da mesma base de dados possibilitada pela mudança entre as diferentes perspectivas [Cano 2007].

2.6. SpagoBI

É uma suíte completa que oferece suporte a negócios cotidianos e estratégicos, tanto em nível de tomada de decisões quanto em nível operacional. Suporta ainda, *Data Mining*,

permitindo descobrir padrões, processo ETL para carregar e gerir a *Data Warehouse*, OLAP que permite a análise multidimensional dos dados, entre outros.

O modelo analítico está formado por diversos motores analíticos, entre os mais usados estão os Relatórios, Indicadores *Key Performance Indicator* (KPI), Painel de Controle, Gráficos, Análise geográfico e Análise multidimensional de dados OLAP.

Assim, esta ferramenta oferece uma solução de BI flexível que proporciona suporte ao monitoramento, análise e apresentação de dados. Desta forma, toda a escalabilidade necessária para a organização é suportada pelo SpagoBI, em termos de arquitetura, funcionalidades e segurança.

3. Implementação de BI

A estratégia utilizada para o desenvolvimento do DM envolve as principais etapas do modelo proposto por [Kimball e Ross 2002], que iniciou-se com o planejamento do projeto e a elaboração do entendimento das necessidades para implementação de BI na UNIPAMPA, assim como, das técnicas e ferramentas necessárias para o desenvolvimento do trabalho.

Um requisito primordial para a criação de DM é a forma com que os dados são organizados nas bases de dados, para que as consultas sejam relevantes no processo decisório. Isto se torna possível através de modelos multidimensionais [Inmon 1997]. Para [Kimball e Ross 2002], a análise da base de dados transacional, permite que o analista obtenha um melhor entendimento das necessidades e limitações do projeto.

Neste momento, se fez possível definir a forma com que os componentes do projeto serão estruturados. Conforme a arquitetura *Bottom-Up*, a construção do ambiente DW é realizada através da implementação de DM independentes, desta forma o projeto tem seu desenvolvimento de forma evolutiva. Para [Kimball e Ross 2002], esta abordagem possui vantagens como, implementação e retorno rápido com enfoque inicial nos principais negócios. Na Figura 1 são ilustrados os componentes da arquitetura de aplicação de BI.

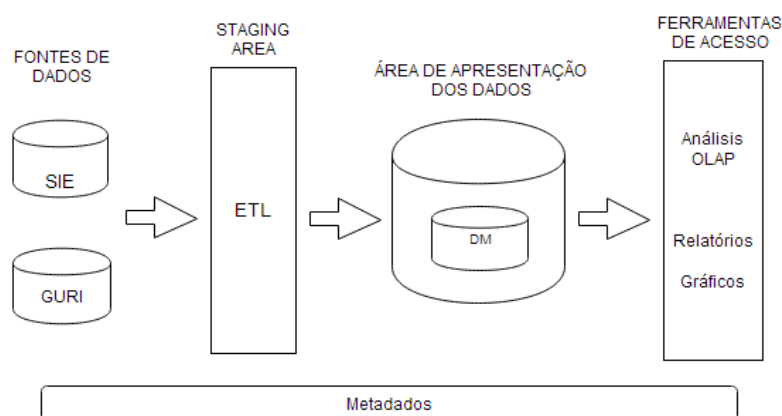


Figura 1. Arquitetura da Aplicação de BI

Com a arquitetura definida, foi iniciado o processo de criação de DM passando por quatro etapas. A primeira delas, consiste na construção do modelo dimensional a partir de um sistema transacional onde é representada a ideia central e suas dimensões, e onde são definidos como os dados serão armazenados, para permitir consultas de maneira

rápida e flexível. A etapa seguinte é o processo ETL, na qual são realizados os processos de extração, transformação e carga de dados de um sistema corporativo para um banco de dados dimensional. E, por fim, a visualização dos resultados a partir da ferramenta para iteração com usuário.

3.1. Modelagem

Para a obtenção dos modelos de dados necessários para o projeto é utilizada a modelagem dimensional. A criação do modelo dimensional no processo de construção de um DM é composto pelos quatro passos, são eles: seleção do processo de negócio, definição da granularidade, escolha das dimensões e escolha dos fatos [Kimball e Ross 2002].

Desta forma, através destas etapas foi construído o modelo dimensional para esta aplicação com auxílio da ferramenta MySQLWorkbench. A Figura 2 apresenta o *Star Schema* da solução.

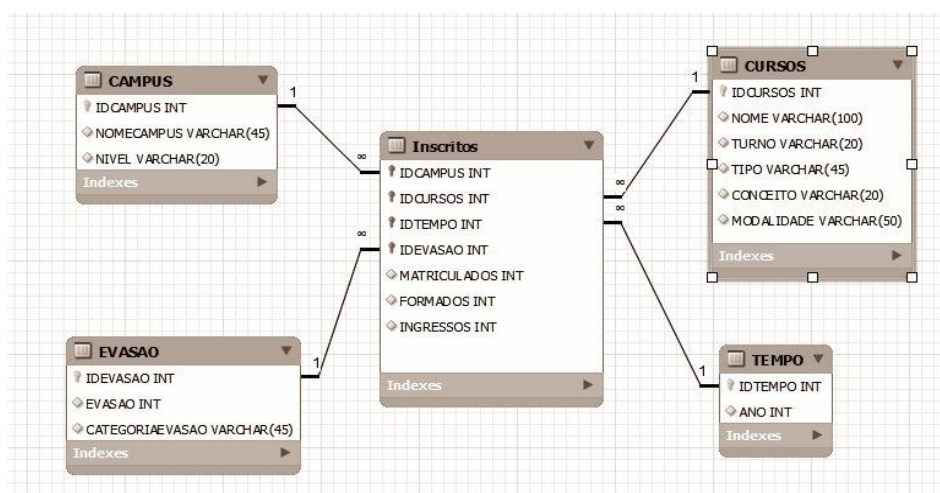


Figura 2. Star Schema

Uma vez realizada a construção do modelo de dados é possível iniciar a etapa de limpeza, transformação e carga dos dados ao repositório. Desta forma, na subseção seguinte é descrito o processo ETL.

3.2. ETL

Para migrar os dados para o DM, o primeiro passo é realizar a extração dos dados fontes, o principal objetivo desta etapa é extrair somente aqueles dados dos sistemas transacionais que serão necessários e prepará-los para as seguintes etapas do processo ETL [Cano 2007].

As bases de dados da UNIPAMPA possuem registros com todas as informações referentes as atividades acadêmicas da Instituição. Realizou-se a coleta de dados utilizados no estudo a partir do Sistema de Informação para o Ensino (SIE) da UNIPAMPA, que possui dados acadêmicos e administrativos da Instituição. Cabe resaltar que para este trabalho a carga de dados foi limitada a um subconjunto de dados destes sistemas.

Na etapa seguinte do processo ETL os dados são pré-processados e preparados

com o objetivo de melhorar a qualidade, gerando uma base separada para a análise. Assim, os dados foram separados e organizados de acordo com as necessidades do projeto. A limpeza contribuiu para eliminar inconsistências da base como, completar dados, tratar valores nulos, e eliminar registros irrelevantes para análise.

Para este trabalho, os dados obtidos foram extraídos e preparados com auxílio da ferramenta Talend, para então serem carregados para o DM apresentando a estrutura necessária para sua utilização.

Na etapa de carga é importante garantir a correspondência entre o relacionamento das tabelas no modelo projetado. Portanto, foi necessário verificar a integridade entre chaves primárias e secundárias, para construir um ambiente analítico íntegro e confiável. Desta maneira, efetuou-se a carga seguindo os modelos de tabelas fatos e dimensões, a partir do script gerado pelo diagrama do modelo construído anteriormente. Desta forma, concluída a etapa de carga dos dados para o DM é possível a visualização dos dados através das ferramentas.

3.3. Exploração dos Dados

Para iniciar o processo de análise e exploração dos dados, foram realizadas a geração de cubo OLAP para possibilitar as consultas, permitindo visões multidimensionais, a criação de gráficos e de relatórios sobre as consultas.

Para a análise OLAP os cubos foram gerados através do SpagoBI e construídos a partir das definições das tabelas fato e de dimensões. Com a estrutura básica do cubo projetada, e a modelagem dimensional previamente descrita, foram definidas para cada dimensão os atributos que seriam visualizados pelos usuários, assim como, suas hierarquias.

Na fase seguinte do processo, foi realizada a visualização dos dados tornando as aplicações de suporte à decisão entendíveis e permitindo uma melhor interpretação dos dados, para possibilitar a identificação de padrões e tendências.

Desta forma, a técnica OLAP permitiu a construção dos gráficos do sistema, possibilitando a visualização das informações estruturadas a partir dos dados existentes no sistema. As operações OLAP realizadas incluem *Slice and Dice*, visualizando dados sob diferentes pontos de vista, aplicando diferentes critérios de filtragem, *Roll Up* agregando informação detalhada a níveis superiores, e *Drill Down* possibilitando o detalhamento dos dados em níveis mais baixos.

3.4. Projeto de Dashboards

Dashboards proporcionam um mecanismo unificador que permite uma base para uma gestão eficaz e eficiente para os projetos. Proporcionam representações gráficas com as informações de negócio de maior relevância para alcançar os objetivos do analista de negócios.

A plataforma SpagoBI possui a tecnologia necessária para a construção dos painéis de controle. Desta maneira, a construção do *dashboard* realizou-se a partir do SpagoBI Studio mostrando a informação de forma útil, resumida e clara.

4. Resultados

A fim de limitar o escopo do problema e viabilizar a concentração de esforços para

conduzir a resultados significativos dentro da realidade da UNIPAMPA, e de sua atividade finalística, foi decidida a implantação de um DM acadêmico. Pode ser observado, que apesar de todos os esforços despendidos dentro da instituição, a UNIPAMPA ainda não dispõe de recursos efetivos para suporte ao processo de tomada de decisão.

No âmbito de interesse dos gestores da instituição estão as questões relacionadas ao desempenho acadêmico e evasão dos alunos em relação aos cursos. Com isto, a fim de obter uma melhor visão sobre questões relacionadas aos cursos de graduação e pós-graduação, foi realizada a construção do primeiro DM, o dm_unipampa. Este DM permitirá realizar consultas como: Avaliações sobre os cursos de graduação e pós-graduação; Análise dos cursos com maiores índices de evasão; Análise sobre número de formandos em todos os campus; Análises com o número de matrículas por campus, podendo verificar aumentos ou diminuições em relação aos últimos anos.

O processo de descoberta de conhecimento a partir de análise de dados tem por finalidade identificar conhecimentos novos e relevantes. Por outro lado, dado o conjunto de dados restrito que foi utilizado neste estudo de caso, as análises apenas apresentam padrões estatísticos, não tendo sido possível a geração de novo conhecimento.

Assim sendo, com a implementação do DM e utilização da ferramenta de *Business Intelligence* foi possível construir uma base de suporte à decisão sob uma das áreas de interesse da UNIPAMPA. Logo, para as primeiras análises construídas, foram obtidos e avaliados os resultados. Nas subseções seguintes são apresentadas algumas das análises realizadas.

4.1. Gráfico Matrículas 2012 e 2013

A Figura 3, a seguir, apresenta o gráfico construído utilizando-se os dados dos alunos matriculados nos anos de 2012 e 2013 referente a todos os campus da UNIPAMPA.

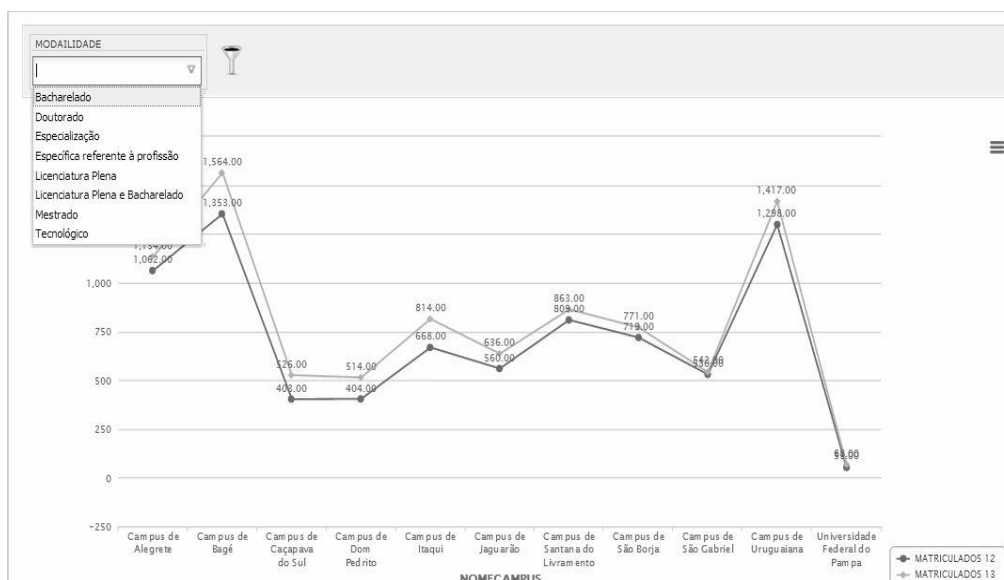


Figura 3. Gráfico Matrículas 2012 e 2013

Com o gráfico, é possível visualizar o aumento no número de matrículas em todos os campus de 2012 para 2013. Filtrando por modalidade, foi possível observar que em relação aos Bacharelados o campus Bagé foi o que obteve maior aumento no número de

matrículas. Enquanto que em relação ao número de matriculados em mestrados, foi constatada uma diminuição no campus Alegrete e um aumento considerável no campus Caçapava do Sul.

4.2. Formandos 2012

Com o gráfico apresentado na Figura 4, a seguir, é possível verificar que o Campus Uruguaiana apresentou o maior número de alunos formados em 2012 em cursos de graduação e pós-graduação, seguido de Bagé e São Borja. Neste mesmo ano, o campus Caçapava do Sul obteve o menor número de alunos formados nestes cursos.

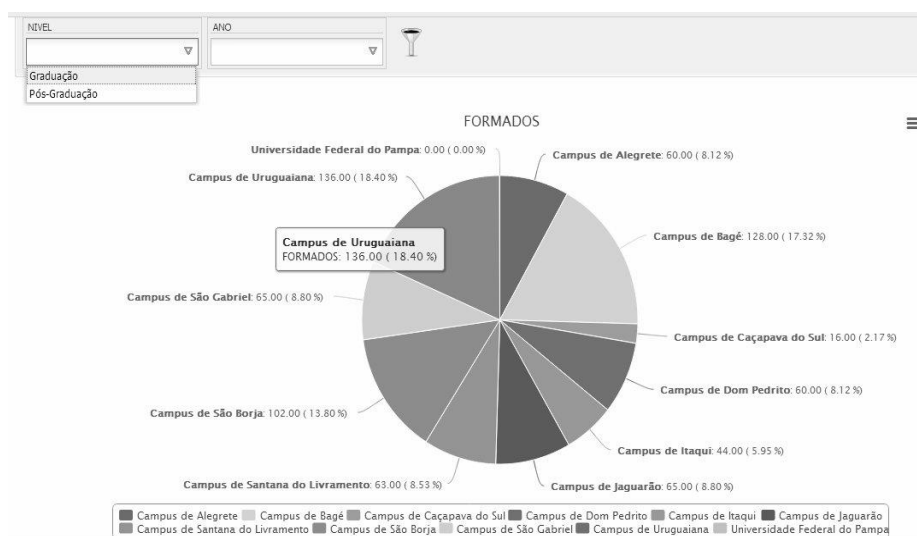


Figura 4. Gráfico Formandos 2012

Com o gráfico é possível verificar que o Campus Uruguaiana apresentou o maior número de alunos formados em 2012 em cursos de graduação e pós-graduação, seguido de Bagé e São Borja. Neste mesmo ano, o campus Caçapava do Sul obteve o menor número de alunos formados nestes cursos.

4.3. Dashboard

4.4. Os *dashboards* permitem uma apresentação visual das informações mais importantes, possibilitando ao usuário avaliações e análises necessárias para alcançar os objetivos de negócio. Na Figura 5 apresenta-se um exemplo de *Dashboard* construído com base nos atributos números de alunos inscritos, número de alunos formados por campus e número de alunos evadidos de cada curso da UNIPAMPA.

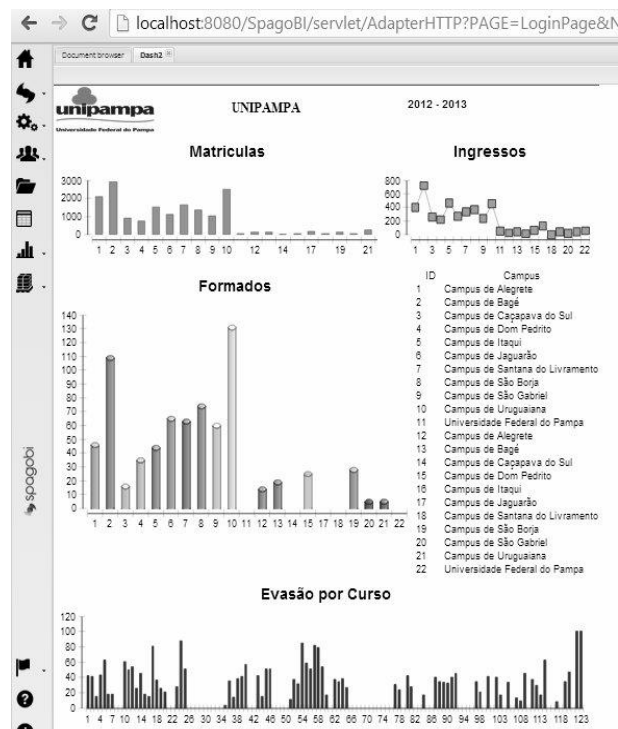


Figura 5. Dashboard UNIPAMPA

Com os resultados obtidos, é possível verificar a importância da implementação de BI na UNIPAMPA. Este tipo de tecnologia possibilita a otimização da entrega de informação, de forma completa, correta, consistente, oportuna e acessível. Facilita o processo de tomada de decisão, oferecendo um maior suporte à informação, possibilitando decisões mais rápidas e maior entendimento do impacto destas decisões. Com isto, as direções de pesquisas futuras incluem questões como: adição de informações ao modelo já construído, afim de mantê-lo atualizado para brindar mais informações e construção de outros *dashboards* para mostrar outros aspectos da DM com dados acadêmicos.

5. Referências

- Cano, J. L. (2007) "Business Intelligence: Competir con Información". Barcelona.
- Diaz, J. C. e Caralt, J. C. (2012) "Introducción al Bussines Intelligence". Editorial UOC. p.17-28, 51-93. Barcelona.
- Inmon, W. H. Como Construir o Data Warehouse. Rio de Janeiro, 1997.
- Júnior, M. C. (2004) "Projetando Sistemas de Apoio à Decisão Baseados em Data Warehouse". 1. ed. Rio de Janeiro: Axcel Books, v. 1.
- Kimball, R. e Ross, M. (2002), The Data Warehouse Toolkit. "The Complete Guide to Dimensional Modeling". 2.ed. John Wiley and Sons Inc.
- Turban, E. et. al. (2009) "Business Intelligence: Um enfoque gerencial para a inteligência do negócio". Bookman.
- Wu, L. et. al. (2007) "A Service-oriented Architecture for Business Intelligence". In: IEEE International Conference on Service-Oriented Computing and Applications(SOCA'07).