

Utilização de Business Intelligence para análise de evasão escolar nos diferentes níveis de ensino do Instituto Federal Farroupilha Campus São Vicente do Sul

Filipe Kulinski Mello¹, Eliana Zen¹, Maicon de Brito do Amarante¹

¹Instituto Federal Farroupilha – Campus São Vicente do Sul (IFFarroupilha SVS)
CEP 97420-000 – São Vicente do Sul– RS – Brasil.

filipe-kulinski@hotmail.com, {maicon.amarante, eliana.zen}
@iffarroupilha.edu.br

Abstract. *In the São Vicente do Sul Campus of the Farroupilha Federal Institute, the school dropouts is challenging, being harmful to the life of the student and, as well for the institution. The first step in treating this problem is collecting information on the current situation. With this in mind, this study aims to use Business Intelligence tools and techniques in the investigation of school dropout in the São Vicente do Sul Campus. This information are shown through charts and tables showing historical data and trends, easing the visualization of situations that present more risk of dropout.*

Resumo. *No Campus São Vicente do Sul do Instituto Federal Farroupilha, a evasão escolar é desafiadora, sendo prejudicial tanto para a vida do aluno como para a vida da instituição. O primeiro passo para o tratamento desse problema é o levantamento de informações sobre a situação atual. Levando em conta essa necessidade, este estudo tem como objetivo utilizar ferramentas e técnicas de Business Intelligence na investigação da evasão escolar do Campus São Vicente do Sul. Essas informações serão mostradas através de gráficos e tabelas apresentando dados históricos e tendências, facilitando a visualização das situações que mais apresentam risco de evasão.*

1. Introdução

O planejamento é uma das atividades de maior complexidade para o gestor. Compreender o cenário atual e pensar no futuro são duas tarefas desafiadoras por si só. Fazer isso sem informações estratégicas que apoiem a decisão pode se tornar inviável. A evasão escolar no Campus São Vicente do Sul do Instituto Federal Farroupilha ilustra precisamente o dilema da decisão. De onde obter informações para embasar a tomada de decisão? Um sistema de gestão acadêmica centralizado, que cubra as diversas áreas institucionais, com dados corretamente atualizados, parece uma solução tentadora. No entanto, chegar a este nível de maturidade pode levar muito tempo, talvez um tempo precioso demais para ficar esperando. Deste modo, como utilizar as informações existentes, ou ainda, apenas aprimorar algumas coletas pontuais de dados, para desde já gerar informações estratégicas, que subsidiem a tomada de decisão?

É neste cenário que surge com força a inteligência de negócios, do inglês *Business Intelligence*, ou simplesmente BI. A implantação de uma ferramenta de BI passa pelas fases de extração, transformação, carga e geração de conhecimento. Durante a fase de extração, diferentes bases de dados, como planilhas, documentos e banco de dados, são extraídas para uma plataforma uniforme. Posteriormente, durante a transformação, dados redundantes e irrelevantes são descartados e informações que não possuíam nenhuma ligação são integradas. Finalmente estes dados são carregados e

organizados para uma *Data Warehouse*, ou depósito de dados, onde as informações são consolidadas e centralizadas (Kimball e Caserta, 2004).

Neste sentido, este projeto tem como objetivo analisar a evasão a partir da compreensão da realidade dos alunos, suas dificuldades e seu grau de satisfação com a instituição, permitindo que assim crie-se um plano de ação que vise reduzir a evasão escolar. As ferramentas e técnicas de BI são um meio concreto para compreendermos nossa realidade, ao passo que entrega uma importante ferramenta de gestão que dará apoio à tomada de decisão.

2. Referencial Teórico

Com o objetivo de fornecer ferramentas que sustentarão a compreensão deste trabalho, esse capítulo visa expor de forma objetiva a fundamentação teórica relacionada ao tema. Serão contextualizados assuntos inerentes ao objeto de pesquisa, tais como: *Business Intelligence*, *On-line Analytical Processing* e *Data Warehouse*.

2.1 Business Intelligence

Business Intelligence, ou simplesmente BI, é um conjunto de conceitos e metodologias de gestão que, fazendo uso de dados extraídos de uma organização, proporciona ganhos nos processos decisórios gerenciais. Baseia-se na capacidade analítica de ferramentas que integram todas as informações necessárias ao processo decisório. O objetivo do *Business Intelligence* é extrair dados estruturados de base de dados, planilhas do Excel, documentos de texto etc, transformá-los e carregá-los em uma *Data Warehouse*, processo esse denominado ETL (*Extraction, Transformation, Loading*). A partir dos dados carregados na *Data Warehouse* é possível gerar Relatórios, *Dashboards* e cubos OLAP que suportem o processo decisório e gere vantagens competitivas (Delsosto, 2014). A Figura 1 exemplifica o processo de *Business Intelligence* descrito acima.

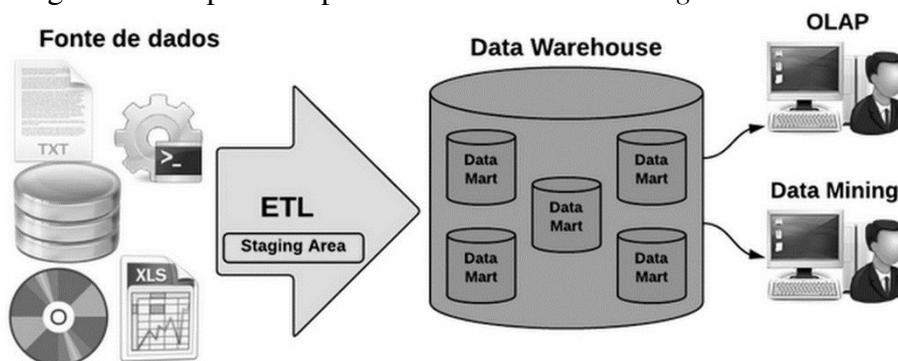


Figura 10. Funcionamento básico do business intelligence

O termo *Business Intelligence* foi primeiramente utilizado pela empresa de consultoria na área de sistemas da informação Gartner Group. Porém, a origem do conceito se iniciou nos anos 70, com base nos sistemas de geração de relatórios de Sistemas de Informação Gerencial (SIG). Segundo Leme Filho (2004), BI é um conjunto de serviços aplicações e tecnologias combinadas para agregar valor, gerenciar e analisar informações. Diante dessa premissa o ambiente de *Business Intelligence* deve possuir cinco características básicas, quais sejam: (a) Extrair e integrar dados de múltiplas fontes; (b) Fazer uso da experiência, democratizando o capital intelectual; (c) Analisar informações contextualizadas, num nível de totalização e agrupamento maior; (e) Identificar relações de causa e efeito; e (d) Desenhar cenários, criar simulações e estudar tendências.

2.2 Data warehouse

Data Warehouse, ou em português armazém de dados, é um depósito de dados único, gerado a partir de diversas fontes de dados, que contém somente informações consideradas importantes para a empresa, ajudando a apoiar as decisões gerenciais. É considerado o principal elemento do BI.

A meta principal de um *Data Warehouse* é a criação de uma visualização única dos dados que residem em diversos bancos de dados físicos, fornecendo aos usuários um modelo de trabalho dos dados da empresa. O acesso a esses dados melhora a qualidade dos serviços e o atendimento ao cliente, ajudando a companhia a avaliar atividades emergentes do negócio (Gonçalves, 2003). O *Data Warehouse* é considerado o principal elemento do BI.

2.2.1 ETL (Extraction, Transformation and Loading)

Segundo Sezões, Oliveira e Baptista (2006), ETL é um conjunto de processos que permite às organizações extrair dados de fontes de informação diversas e reformulá-los e carregá-los para uma nova aplicação (base de dados, geralmente um *Data Warehouse*) para análise.

A ideia principal do ETL é coletar dados de diferentes fontes, mas que sejam pertinentes a uma única regra de negócio. Segundo Kimball e Caserta (2004), o processo de ETL divide-se em três etapas principais: (1) Etapa de extração, onde são coletados dados de diferentes tipos e sistemas de origem; (2) Etapa de transformação, onde são aplicadas séries de regras e funções que padronizam o formato dos dados extraídos; (3) Fase de Carga, onde os dados transformados são transferidos para os locais de destino, como *Data Warehouse* e *Data Marts*.

2.3 Cubos OLAP (On-line Analytical Processing)

A tecnologia OLAP representa a possibilidade de se trabalhar os dados, com operadores dimensionais, possibilitando uma forma múltipla e combinada de análise (Barbieri, 2001).

A característica principal dos sistemas OLAP é permitir uma visão conceitual multidimensional dos dados armazenados, através de cubos (Figura 3). Um cubo é uma estrutura composta de dimensões e de uma tabela fato. As dimensões representam os eixos do cubo e correspondem a atributos do domínio analisado. A tabela fato é constituída de medidas (dados) de uma tabela de fatos coletados dentro do domínio analisado. As medidas são relacionadas às dimensões (Tronto e Sant'anna, 2004).

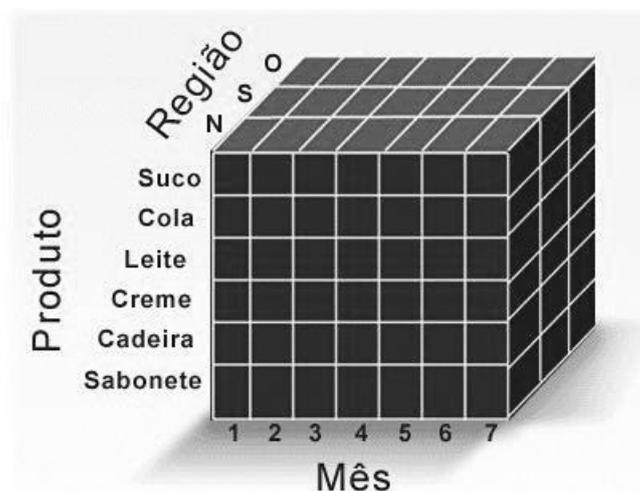


Figura 2 Estrutura de um cubo OLAP

Com a utilização de cubos OLAP, é possível visualizar determinados cubos de informações de diferentes ângulos, e de vários níveis de agregação, permitindo a análise de um grande volume de dados em tempo real e obtenção de relatórios analíticos. É forma rápida de obtenção de informações gerenciais para a tomada de decisão dos gestores.

3. Materiais e métodos

Para a implementação do projeto proposto foram utilizadas apenas ferramentas gratuitas e *open source*, das quais pode-se citar: (a) Pentaho BI Server 5.0.1, responsável por toda a camada servidora da plataforma do Pentaho; (b) *plug-ins* Saiku, Pentaho Dashboards Editor e Pentaho Data Access, utilizados para a visualização de *dashboards* e Cubos OLAP; e (c) Pentaho Report Designer, para a geração de Relatórios Estatísticos. Para o desenvolvimento de cubos OLAP foram utilizadas as ferramentas Mondrian Workbench, Pentaho Data Integration (Kettle) e o Sistema Gerenciado de Banco de Dados PostgreSQL.

3.1 Pentaho

A plataforma Pentaho possui um servidor web que disponibiliza um conjunto de serviços, dos quais destaca-se a autenticação de usuários, relatórios, análises exploratórias, *dashboards*, dentre outras. A plataforma Pentaho BI é constituída por dois servidores web, um para acomodar a aplicação do usuário final, chamada Pentaho User Console (BI-Server-CE), e o outro para acomodar a aplicação do Administrador, chamada Pentaho Administration Console.

A utilização da ferramenta Pentaho Bi-server mostrou-se eficaz no desenvolvimento da pesquisa, devido a facilidade de se construir e acessar todas as funcionalidades e resultados gerados dinamicamente pela ferramenta por meio de uma interface web amigável e intuitiva, destacando também o fato de ser uma ferramenta gratuita, com excelente documentação e referências bibliográficas.

4. Metodologia de pesquisa

A presente pesquisa pode ser definida como quantitativa tradicional, já que busca apresentar fatos através da descrição e interpretação de fatos isolados. Segundo Berto e Nakano (2000), a pesquisa qualitativa envolve natureza empírica, relações de causa e efeito, hipóteses bem formuladas e métodos lógico-dedutivos que permitem replicar resultados através da generalização.

A pesquisa realizada levou em consideração a base de dados do sistema acadêmico do Instituto Federal Farroupilha Campus São Vicente do Sul e considerou dados do período compreendido entre janeiro de 2010 e dezembro de 2014. O processo de pesquisa foi dividido em quatro etapas: (1) estudo do estado da arte; (2) coleta análise e processamento de dados; (3) construção de gráficos, compilação de informações estratégicas; e (4) análise e avaliação dos resultados. A Tabela 1 descreve melhor cada uma das etapas envolvidas no processo de pesquisa.

Etapa	Descrição
1	Visou obter uma melhor compreensão do problema, através do estudo do estado da arte e das ferramentas empregadas durante a pesquisa.
2	Utilizou-se a ferramenta Data Integration Kettle da suíte Pentaho para coletar dados relevantes do sistema acadêmico do Campus, os quais, posteriormente, foram carregados em uma <i>Data Warehouse</i> ,
3	Envolveu a compilação das informações de acordo com os dados coletados da fase anterior. Foram utilizadas ferramentas da suíte Pentaho para gerar relatórios, <i>dashboards</i> e um Cubo OLAP apresentando informações

	relacionadas a evasão escolar proporcional ao número de alunos matriculados.
4	Realizou-se a análise dos dados, avaliando a relação de evasão/matrícula, buscando estabelecer relações e correlações de causa e efeito, procurando discrepâncias e refinando a apresentação dos resultados.

Tabela 1. Etapas do processo de pesquisa

A análise dos dados foi dividida em três dimensões: de ensino, de espaço e de tempo. A dimensão de ensino apresenta a evasão dos alunos conforme o curso e nível de ensino; a dimensão de espaço mostra os dados conforme a localização do aluno evadido, sendo dividida em evasão por cidade e por distância da cidade de endereço do aluno; e a dimensão tempo, que mostra a evasão por idade e período do curso em que ocorreu a evasão. Neste trabalho é descrita a análise dos dados relativa a dimensão de ensino.

5. Resultados

Na análise de resultados, foi observado os dados de evasão dos cursos e dos níveis de ensino do Campus. A análise levou em consideração o número total de matrículas, o total de evasões e a relação matrícula/evasão, esta última, relacionada ao comparativo entre alunos evadidos e matriculados em determinado período.

5.1 Evasão por nível de ensino

Primeiramente, foram analisados os dados de evasão referentes aos níveis de ensino da instituição. A Tabela 1 apresenta o número de matrículas, evasões e relação matrícula/evasão.

	Ano	Matrículas	Evasões	Relação Matrícula / Evasão
Nível Médio	2010	351	13	3.70 %
	2011	405	33	8.15 %
	2012	481	45	9.36 %
	2013	388	59	15.21 %
	2014	497	94	18.91 %
Nível Técnico	2010	1279	47	3.67 %
	2011	1262	119	9.43 %
	2012	1080	117	10.83 %
	2013	764	114	14.92 %
	2014	624	124	19.87 %
Nível Superior	2010	608	29	4.77 %
	2011	818	69	8.44 %
	2012	900	97	10.78 %
	2013	834	100	11.99 %
	2014	736	124	16.85 %

Tabela 2. Evasão por nível de ensino

Com relação ao nível médio é possível observar que houve um aumento de 5.86% no percentual de evasão e uma queda de 107 alunos matriculados no período de 2013. No ano seguinte, o número de matrículas aumentou em 109 enquanto a relação matrícula/evasão subiu proporcionalmente ao número de alunos.

Quanto ao nível superior, também houve uma queda de 66 matrículas em 2013 juntamente com um aumento de 1,21 % no percentual de evasão em relação a 2012. Porém, diferentemente do nível médio, o número de matrículas continuou a cair em 2014, havendo uma diminuição de 98 alunos matriculados e um aumento de 4.86 % na relação matrícula/evasão.

A maior queda de matrículas em 2013 ocorreu no nível técnico, pois houve uma diminuição de 316 matrículas e um aumento de 4,09 % na relação matrícula/evasão. A

queda de matrículas continuou ocorrendo no ano seguinte, havendo uma diminuição de 140 matrículas e aumento de 4,95 % na relação matrícula/evasão.

Ao analisar os dados, pode-se considerar que o nível técnico possui a situação de evasão mais grave entre os três níveis, pois tem o maior percentual de evasão, além de apresentar uma maior queda de matrículas de 2012 para 2013 e uma diminuição de matrículas entre o período de 2010 a 2014.

5.2 Evasão por curso

A Tabela 2 mostra a relação matrícula/evasão (representada pelo símbolo “%”) dos dez cursos com maior ocorrência de evasão no instituto, ordenados pelo número total de evasões. Analisando essa tabela, é possível observar um aumento considerável na relação evasão/matrícula de todos os cursos no período de 2010 a 2014.

	2010	2011	2012	2013	2014
Curso	%	%	%	%	%
Técnico em Informática	4,05 %	10,86 %	14,51 %	17,03 %	37,36 %
Técnico em Agropecuária interado ao Ensino Médio	4,50 %	7,17 %	6,85 %	14,12 %	17,92 %
Tecnólogo em Análise e Desenvolvimento de Sistemas	3,46 %	6,48 %	18,86 %	14,47 %	24,82 %
Tecnólogo em Gestão Pública	5,91 %	8,65 %	7,72 %	10,48 %	11,79 %
Técnico em Secretariado	5,36 %	15,38 %	11,98 %	22,48 %	30,00 %
Técnico em Agricultura	2,59 %	7,32 %	6,29 %	8,02 %	11,49 %
Técnico em Zootecnia	2,66 %	6,98 %	13,41 %	15,75 %	20,00 %
Licenciatura em Biologia	4,88 %	8,70 %	4,74 %	9,96 %	15,04 %
Licenciatura em Química		6,78 %	17,65 %	14,78 %	22,70 %
Técnico Integrado em Manutenção e Suporte a informática - MSI			15,87 %	20,75 %	21,92 %

Tabela 3. Evasão por Curso

O curso Técnico em Informática foi o que apresentou a situação mais grave de evasão, já que possuiu o maior percentual de evasão no ano de 2014, 37,36%, tendo também o maior aumento em relação a 2013, chegando a 20,33 %. Também pode-se chamar atenção aos cursos Técnico em Secretariado e Tecnólogo em Análise e Desenvolvimento de Sistemas, que também apresentaram um alto percentual de evasão em 2014, além de um aumento no percentual de evasão de 7,53% e 10,35%, respectivamente, em relação a 2013.

Os menores índices de evasão ocorreram nos cursos Tecnólogo em Gestão Pública e Técnico em Agricultura, onde houve 11,79% e 11,49% de evasão no ano de 2014 e um aumento de 1,31% e 3,47%, respectivamente, em relação a 2013. Os cursos que não possuem percentual ainda não haviam sido criados no determinado período.

6. Conclusão

Certamente, a implementação de um sistema de Business Intelligence (BI) é de grande valia para a gerência de uma instituição. As ferramentas oferecidas por essa tecnologia são cada vez mais necessárias nos processos de tomada de decisão. A maior vantagem do BI é permitir acesso a informação de qualidade em um curto período de tempo, permitindo aos gestores conhecerem melhor a realidade da instituição.

Através da implementação do sistema de BI no Campus, foi possível entregar aos gestores uma plataforma de fácil acesso à informação, apresentando um cubo OLAP, gráficos e relatórios que facilitam a consulta a informações e o cruzamento de dados relacionados a evasão escolar.

A partir desse projeto, foi possível entender melhor a situação da evasão no Campus. O estudo apresentou situações em que o aluno tem mais chance de evadir, que até então eram desconhecidas pelos gestores. Através dele foi possível observar o aumento crescente da evasão em todos os níveis de ensino, sendo mais grave nos cursos de nível técnico, juntamente com outras tendências de evasão, como cursos com maiores índices de abandono de alunos. Futuramente, essas informações poderão ser utilizadas na geração de estratégias para mitigar a evasão escolar no Campus São Vicente do Sul.

Referências

- KIMBALL, R e ROSS, M. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Kimball Group, 2013.
- KIMBALL, R e CASERTA, Joe. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data, Kimball Group, 2004.
- LEME FILHO, Trajano. Business Intelligence no Microsoft Excel. Rio de Janeiro, Axcel Books do Brasil, 2004.
- ARRUDA, Cláudio. Estudo para implantação de um DataWarehouse em um ambiente empresarial. Santa Catarina, 2003.
- BARBIERI, Carlos. BI- Business Intelligence: Modelagem & Tecnologia. Rio de Janeiro, Axcel Books do Brasil , 2001.
- TRONTO, I. F. Barcellos e SANT'ANNA, Nilson. Um Roteiro para Construção de Cubos e Consultas OLAP. 2004. Disponível em:<http://mtcm18.sid.inpe.br/col/lac.inpe.br/worcap/2004/10.05.09.36/doc/WorCapQuatroIris.PDF>.
- SEZÕES, Carlos; OLIVEIRA, José; BAPTISTA Miguel. Business Intelligence. Porto: Sociedade Portuguesa de Inovação, 2006.
- BERTO, R. M. V. S., and Davi Noboru NAKANO. "A produção científica nos anais do Encontro Nacional de Engenharia de Produção: um levantamento de métodos e tipos de pesquisa." Revista Produção 9.2 (2000): 65-76.
- KIMBALL, R.; CASERTA, J. The Data Warehouse ETL Toolkit. Indianapolis: Wiley Publishing, Inc, 2004.